

# Multiperiod Martingale Transport

Marcel Nutz\*      Florian Stebegg†      Xiaowei Tan‡

March 31, 2017

## Abstract

Consider a multiperiod optimal transport problem where distributions  $\mu_0, \dots, \mu_n$  are prescribed and a transport corresponds to a scalar martingale  $X$  with marginals  $X_t \sim \mu_t$ . We introduce particular couplings called left-monotone transports; they are characterized equivalently by a no-crossing property of their support, as simultaneous optimizers for a class of bivariate transport cost functions with a Spence–Mirrlees property, and by an order-theoretic minimality property. Left-monotone transports are unique if  $\mu_0$  is atomless, but not in general. In the one-period case  $n = 1$ , these transports reduce to the Left-Curtain coupling of Beiglböck and Juillet. In the multiperiod case, the bivariate marginals for dates  $(0, t)$  are of Left-Curtain type, if and only if  $\mu_0, \dots, \mu_n$  have a specific order property. The general analysis of the transport problem also gives rise to a strong duality result and a description of its polar sets. Finally, we study a variant where the intermediate marginals  $\mu_1, \dots, \mu_{n-1}$  are not prescribed.

*Keywords:* Optimal Transport; Martingale Coupling; Duality

*AMS 2010 Subject Classification:* 60G42; 49N05

## 1 Introduction

Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be a vector of probability measures  $\mu_t$  on the real line. A measure  $P$  on  $\mathbb{R}^{n+1}$  whose marginals are given by  $\boldsymbol{\mu}$  is called a coupling (or transport) of  $\boldsymbol{\mu}$ , and the set of all such measures is denoted by  $\Pi(\boldsymbol{\mu})$ . We shall be interested in couplings  $P$  that are martingales; that is, the identity  $X = (X_0, \dots, X_n)$  on  $\mathbb{R}^{n+1}$  is a martingale under  $P$ . Hence, we will assume that all marginals have a finite first moment and denote by  $\mathcal{M}(\boldsymbol{\mu})$  the set

---

\*Departments of Statistics and Mathematics, Columbia University, mnutz@columbia.edu. Research supported by an Alfred P. Sloan Fellowship and NSF Grant DMS-1512900.

†Department of Statistics, Columbia University, florian.stebegg@columbia.edu.

‡Department of Mathematics, Columbia University, xt2161@columbia.edu.

of martingale couplings. A classical result of Strassen [39] shows that  $\mathcal{M}(\boldsymbol{\mu})$  is nonempty if and only if the marginals are in convex order, denoted by  $\mu_{t-1} \leq_c \mu_t$  and defined by the requirement that  $\mu_{t-1}(\phi) \leq \mu_t(\phi)$  for any convex function  $\phi$ , where  $\mu(\phi) := \int \phi d\mu$ .

The first goal of this paper is to introduce and study a family of “canonical” couplings  $P \in \mathcal{M}(\mu)$  that we call left-monotone. These couplings specialize to the Left-Curtain coupling of [8] in the one-step case  $n = 1$  and share, broadly speaking, several properties reminiscent of the Hoeffding–Fréchet coupling of classical optimal transport. Indeed, left-monotone couplings will be characterized by order-theoretic minimality properties, as simultaneous optimal transports for certain classes of reward (or cost) functions, and through no-crossing conditions on their supports.

The second goal is to develop a strong duality theory for multiperiod martingale optimal transport, along the lines of [10] for the one-period martingale case and [32] for the classical optimal transport problem. That is, we introduce a suitable dual optimization problem and show the absence of a duality gap as well as the existence of dual optimizers for general transport reward (or cost) functions. The duality result is a crucial tool for the study of the left-monotone couplings.

We also develop similar results for a variant of our problem where the intermediate marginals  $\mu_1, \dots, \mu_{n-1}$  are not prescribed (Section 9), but we shall focus on the full marginal case for the purpose of the Introduction.

## 1.1 Left-Monotone Transports

For the sake of orientation, let us first state the main result and then explain the terminology contained therein. The following is a streamlined version—the results in the body of the paper are stronger in some technical aspects.

**Theorem 1.1.** *Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order and  $P \in \mathcal{M}(\boldsymbol{\mu})$  a martingale transport between these marginals. The following are equivalent:*

- (i)  *$P$  is a simultaneous optimal transport for  $f(X_0, X_t)$ ,  $1 \leq t \leq n$  whenever  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is a smooth second-order Spence–Mirrlees function.*
- (ii)  *$P$  is concentrated on a left-monotone set  $\Gamma \subseteq \mathbb{R}^{n+1}$ .*
- (iii)  *$P$  transports  $\mu_0|_{(-\infty, a]}$  to the obstructed shadow  $\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0|_{(-\infty, a]})$  in step  $t$ , for all  $1 \leq t \leq n$  and  $a \in \mathbb{R}$ .*

*There exists  $P \in \mathcal{M}(\boldsymbol{\mu})$  satisfying (i)–(iii), and any such  $P$  is called a left-monotone transport. If  $\mu_0$  is atomless, then  $P$  is unique.*

Let us now discuss the items in the theorem.

**(i) Optimal Transport.** This property characterizes  $P$  as a simultaneous optimal transport. Given a function  $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ , we may consider the martingale optimal transport problem with reward  $f$  (or cost  $-f$ ),

$$\mathbf{S}_\mu(f) = \sup_{P \in \mathcal{M}(\mu)} P(f); \quad (1.1)$$

recall that  $P(f) = \mathbb{E}^P[f(X_0, \dots, X_n)]$ . A Lipschitz function  $f \in C^{1,2}(\mathbb{R}^2; \mathbb{R})$  is called a *smooth second-order Spence–Mirrlees* function if it satisfies the cross-derivative condition  $f_{xyy} > 0$ ; this has also been called the martingale Spence–Mirrlees condition in analogy to the classical Spence–Mirrlees condition  $f_{xy} > 0$ . Given such a function of two variables and  $1 \leq t \leq n$ , we may consider the  $n$ -step martingale optimal transport problem with reward  $f(X_0, X_t)$ . Characterization (i) states that a left-monotone transport  $P \in \mathcal{M}(\mu)$  is an optimizer simultaneously for the  $n$  transport problems  $f(X_0, X_t)$ ,  $1 \leq t \leq n$ , for some (and then all) smooth second-order Spence–Mirrlees functions  $f$ .

In the one-step case, a corresponding result holds for the Left-Curtain coupling [8]; here the simultaneous optimization becomes a single one. In view of the characterization in (i), an immediate consequence is that if there exists  $P \in \mathcal{M}(\mu)$  such that all bivariate projections  $P_{0t} = P \circ (X_0, X_t)^{-1} \in \mathcal{M}(\mu_0, \mu_t)$  are of Left-Curtain type, then  $P$  is left-monotone. However, such a transport does not exist unless the marginals satisfy a very specific condition (see Proposition 6.9), and in general the bivariate projections of a left-monotone transport are *not* of Left-Curtain type.

**(ii) Geometry.** The second item characterizes  $P$  through a geometric property of its support. A set  $\Gamma \subseteq \mathbb{R}^{n+1}$  will be called *left-monotone* if it has the following no-crossing property for all  $1 \leq t \leq n$ : Let  $\mathbf{x} = (x_0, \dots, x_{t-1})$ ,  $\mathbf{x}' = (x'_0, \dots, x'_{t-1}) \in \mathbb{R}^t$  and

$$y^-, y^+, y' \in \mathbb{R} \text{ with } y^- < y^+$$

be such that  $(\mathbf{x}, y^+)$ ,  $(\mathbf{x}, y^-)$ ,  $(\mathbf{x}', y')$  are in the projection of  $\Gamma$  to the first  $t+1$  coordinates. Then,

$$y' \notin (y^-, y^+) \text{ whenever } x_0 < x'_0.$$

That is, if we consider two paths in  $\Gamma$  starting at  $x_0$  and coinciding up to  $t-1$ , and a third path starting at  $x'_0$  to the right of  $x_0$ , then at time  $t$  the third path cannot step in-between the first two—this is illustrated in Figure 1. Item (ii) states that a left-monotone transport  $P \in \mathcal{M}(\mu)$  can be

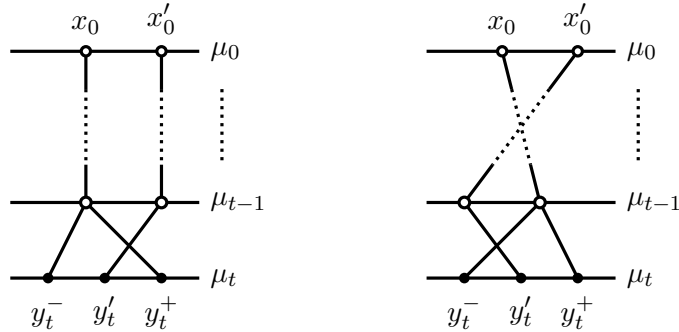


Figure 1: Two examples of forbidden configurations in left-monotone sets.

characterized by the fact that it is concentrated on a left-monotone set  $\Gamma$ . (In Theorem 7.16 we shall state a stronger result: we can find a left-monotone set that carries all left-monotone transports at once.)

In the one-step case  $n = 1$ , left-monotonicity coincides with the Left-Curtain property of [8]. However, we emphasize that for  $t > 1$ , our no-crossing condition differs from the Left-Curtain property of the bivariate projection  $(X_0, X_t)(\Gamma)$  as the latter would not contain the restriction that the first two paths have to coincide up to  $t - 1$  (see also Example 6.10). This corresponds to the mentioned fact that the bivariate marginal  $P_{0t}$  need not be of Left-Curtain type. On the other hand, the geometry of the projection  $(X_{t-1}, X_t)(\Gamma)$  is also quite different from the Left-Curtain one, as our condition may rule out third paths crossing from the right *and* left at  $t - 1$ , depending on the starting point  $x'_0$  rather than the location of  $x'_{t-1}$ .

**(iii) Convex Ordering.** This property characterizes left-monotone transports in an order-theoretic way and will be used in the existence proof. To explain the idea, suppose that  $\mu_0$  consists of finitely many atoms at  $x_1, \dots, x_N \in \mathbb{R}$ . Then, for any fixed  $t$ , a coupling of  $\mu_0$  and  $\mu_t$  can be defined by specifying a “destination” measure for each atom. We consider all chains  $\mu_0|_{x_i} \leq_c \theta_1 \leq_c \dots \leq_c \theta_t$  of measures  $\theta_s$  in convex order that satisfy the marginals constraints  $\theta_s \leq \mu_s$  for  $s \leq t$ . Of these chains, keep only the terminal measures  $\theta_t$  and rank them according to the convex order. The *obstructed shadow of  $\mu_0|_{x_i}$  in  $\mu_t$  through  $\mu_1, \dots, \mu_{t-1}$* , denoted  $\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0|_{x_i})$ , is defined as the unique least element among the  $\theta_t$ . A particular coupling of  $\mu_0$  and  $\mu_t$  is the one that successively maps the atoms  $\mu_0|_{x_i}$  to their obstructed shadows, starting with the left-most atom  $x_i$  and continuing from

left to right. In the case of general measures, we consider the restrictions  $\mu_0|_{(-\infty, a]}$  instead of successively mapping the atoms. Characterization (iii) then states that a left-monotone transport  $P \in \mathcal{M}(\boldsymbol{\mu})$  maps  $\mu_0|_{(-\infty, a]}$  to its obstructed shadow at date  $t$  for all  $1 \leq t \leq n$  and  $a \in \mathbb{R}$ . This shows in particular that the bivariate projections  $P_{0t} = P \circ (X_0, X_t)^{-1}$  of a left-monotone coupling are uniquely determined. In the body of the text, we shall also give an alternative definition of the obstructed shadow by iterating unobstructed shadows through the marginals up to date  $t$ ; see Section 6.

The above specializes to the construction of [8] for the one-step case, which corresponds to the situation of  $t = 1$  where there are no intermediate marginals obstructing the shadow. When  $t > 1$ , the obstruction by the intermediate marginals once again entails that  $P_{0t}$  need not be of Left-Curtain type. More precisely, Characterization (iii) gives rise to a sharp criterion (Proposition 6.9) on the marginals  $\boldsymbol{\mu}$ , describing exactly when this coincidence arises.

**(Non-)Uniqueness.** We have seen above that for a left-monotone transport  $P \in \mathcal{M}(\boldsymbol{\mu})$  the bivariate projections  $P_{0t}$ ,  $1 \leq t \leq n$  are uniquely determined. In particular, for  $n = 1$ , we recover the result of [8] that the Left-Curtain coupling is unique. For  $n > 1$ , the situation turns out to be quite different depending on the nature of the first marginal. On the one extreme, we shall see that when  $\mu_0$  is atomless, there is a unique left-monotone transport  $P \in \mathcal{M}(\boldsymbol{\mu})$ . Moreover,  $P$  has a degenerate structure reminiscent of Brenier’s theorem: it can be disintegrated as  $P = \mu_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_n$  where each one-step transport kernel  $\kappa_t$  is concentrated on the graphs of two functions. On the other extreme, if  $\mu_0$  is a Dirac mass, the typical case is that there are infinitely many left-monotone couplings—see Section 8 for a detailed discussion. We shall also show that left-monotone transports are not Markovian in general, even if uniqueness holds (Example 7.17).

## 1.2 Duality

The analysis of left-monotone transports is based on a duality result that we develop for general reward functions  $f : \mathbb{R}^{n+1} \rightarrow (-\infty, \infty]$  with an integrable lower bound. Formally, the dual problem (in the sense of linear programming) for the transport problem  $\mathbf{S}_{\boldsymbol{\mu}}(f) = \sup_{P \in \mathcal{M}(\boldsymbol{\mu})} P(f)$  is the minimization

$$\mathbf{I}_{\boldsymbol{\mu}}(f) := \inf_{(\phi, H)} \sum_{t=0}^n \mu_t(\phi_t)$$

where the infimum is taken over vectors  $\phi = (\phi_0, \dots, \phi_n)$  of real functions and predictable processes  $H = (H_1, \dots, H_n)$  such that

$$\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \geq f; \quad (1.2)$$

here  $(H \cdot X)_n := \sum_{t=1}^n H_t(X_t - X_{t-1})$  is the discrete-time integral. The desired result (Theorem 5.2) states that there is no duality gap, i.e.  $\mathbf{I}_\mu(f) = \mathbf{S}_\mu(f)$ , and that the dual problem is attained whenever it is finite. From the analysis for the one-step case in [10] we know that this assertion fails for the above naive formulation of the dual, and requires several relaxations regarding the integrability of the functions  $\phi_t$  and the domain  $\mathcal{V} \subseteq \mathbb{R}^{n+1}$  where the inequality (1.2) is required. Specifically, the inequality needs to be relaxed on sets that are  $\mathcal{M}(\mu)$ -polar; i.e. not charged by any transport  $P \in \mathcal{M}(\mu)$ . These sets are characterized in Theorem 3.1 where we show that the  $\mathcal{M}(\mu)$ -polar sets are precisely the (unions of) sets which project to a two-dimensional polar set of  $\mathcal{M}(\mu_{t-1}, \mu_t)$  for some  $1 \leq t \leq n$ .

The duality theorem gives rise to a monotonicity principle (Theorem 5.4) that underpins the analysis of the of left-monotone couplings. Similarly to the cyclical monotonicity condition in classical transport, it allows one to study the geometry of the support of optimal transports for a given function  $f$ .

### 1.3 Background and Related Literature

The martingale optimal transport problem (1.1) was introduced in [5] with the dual problem as a motivation. Indeed, in financial mathematics the function  $f$  is understood as the payoff of a derivative written on the underlying  $X$  and (1.2) corresponds to superhedging  $f$  by statically trading in European options  $\phi_t(X_t)$  and dynamically trading in the underlying according to the strategy  $H$ . The value  $\mathbf{I}_\mu(f)$  then corresponds to the lowest price of  $f$  for which the seller can enter a model-free hedge  $(\phi, H)$  if the marginals  $X_t \sim \mu_t$  are known from option market data. In [5], it was shown (with the above, “naive” formulation of the dual problem) that there is no duality gap if  $f$  is sufficiently regular, whereas dual existence was shown to fail even in regular cases. The idea of model-free hedging as well as the connection to Skorokhod embeddings goes back to [24]; we refer to [12, 13, 14, 25, 36, 41] for further references. A specific multiperiod martingale optimal transport problem also arises in the study of the maximum maximum of a martingale given  $n$  marginals [20].

The one-step case  $n = 1$  has been studied in great detail. In particular, [8] introduced the Left-Curtain coupling and pioneered numerous ideas underlying Theorem 1.1, [22] provided an explicit construction of that coupling, and [30] established the stability with respect to the marginals. Our duality results specialize to the ones of [10] when  $n = 1$ . Unsurprisingly, we shall exploit many arguments and results from these papers wherever possible. As indicated above, and as will be seen in the proofs below, the multistep case allows for a richer structure and necessitates novel ideas; for instance, the analysis of the polar sets (Theorem 3.1) is surprisingly involved. Other works in the one-step martingale case have studied reward functions  $f$  such as forward start straddles [26, 27] or Asian payoffs [38]. We also refer to [18, 33] for recent developments with multidimensional marginals.

One-step martingale optimal transport problems can alternately be studied as optimal Skorokhod embedding problems with marginal constraints; cf. [2, 3, 6, 7]. A multi-marginal extension [1] of [2] is in preparation at the time of writing and the authors have brought to our attention that it will offer a version of Theorem 1.1 in the Skorokhod picture, at least in the case where  $\mu_0$  is atomless and some further conditions are satisfied. The Skorokhod embedding problem with multi-marginal constraint was also studied in [19].

A multi-step coupling quite different from ours can be obtained by composing in a Markovian fashion the Left-Curtain transport kernels from  $\mu_{t-1}$  to  $\mu_t$ ,  $1 \leq t \leq n$ , as discussed in [22]. In [29] the continuous-time limits of such couplings for  $n \rightarrow \infty$  are studied to find solutions of the so-called Peacock problem [23] where the marginals for a continuous-time martingale are prescribed; see also [21] and [31] for other continuous-time results with full marginal constraint. Early contributions related to the continuous-time martingale transport problem include [15, 16, 17, 34, 37, 40].

The remainder of the paper is organized as follows. Section 2 fixes basic terminology and recalls the necessary results from the one-step case. In Section 3, we characterize the polar structure of  $\mathcal{M}(\mu)$ . Section 4 introduces and analyzes the space that is the domain of the dual problem in Section 5, where we state the duality theorem and the monotonicity principle. Section 6 introduces left-monotone transports by the shadow construction and Section 7 develops the equivalent characterizations in terms of support and optimality properties. The (non-)uniqueness of left-monotone transports is discussed in Section 8. We conclude with the analysis of the problem with unconstrained intermediate marginals in Section 9.

## 2 Preliminaries

Throughout this paper,  $\mu_t, \mu, \nu$  denote finite measures on  $\mathbb{R}$  with finite first moment, the total mass not necessarily being normalized. Generalizing the notation from the Introduction to a vector  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  of such measures, we will write  $\Pi(\boldsymbol{\mu})$  for the set of couplings; that is, measures  $P$  on  $\mathbb{R}^{n+1}$  such that  $P \circ X_t^{-1} = \mu_t$  for  $0 \leq t \leq n$  where  $X = (X_0, \dots, X_n) : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$  is the identity. Moreover,  $\mathcal{M}(\boldsymbol{\mu})$  is the subset of all  $P \in \Pi(\boldsymbol{\mu})$  that are martingales, meaning that

$$\int X_s \mathbf{1}_A(X_0, \dots, X_s) dP = \int X_t \mathbf{1}_A(X_0, \dots, X_s) dP$$

for all  $s \leq t$  and Borel sets  $A \in \mathfrak{B}(\mathbb{R}^{s+1})$ .

We denote by  $\mathbb{F} = \{\mathfrak{F}_t\}_{0 \leq t \leq n}$  the canonical filtration  $\mathfrak{F}_t := \sigma(X_0, \dots, X_t)$ . As usual, an  $\mathbb{F}$ -predictable process  $H = \{H_t\}_{1 \leq t \leq n}$  is a sequence of real functions on  $\mathbb{R}^{n+1}$  such that  $H_t$  is  $\mathfrak{F}_{t-1}$ -measurable; i.e.  $H_t = h_t(X_0, \dots, X_{t-1})$  for some Borel-measurable  $h_t : \mathbb{R}^t \rightarrow \mathbb{R}$ . Given an  $\mathbb{F}$ -predictable process  $H$ , the discrete stochastic integral  $\{(H \cdot X)_t\}_{0 \leq t \leq n}$  is defined by

$$(H \cdot X)_t := \sum_{s=1}^t H_s \cdot (X_s - X_{s-1}).$$

If  $X$  is a martingale under some measure  $P$ , then  $H \cdot X$  is a generalized (not necessarily integrable) martingale in the sense of generalized conditional expectations; cf. [28, Proposition 1.64].

We say that  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  is in convex order if  $\mu_{t-1} \leq_c \mu_t$  for all  $1 \leq t \leq n$ ; that is,  $\mu_{t-1}(\phi) \leq \mu_t(\phi)$  for any convex function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$ . This implies that  $\mu_{t-1}$  and  $\mu_t$  have the same total mass. The order can also be characterized by the potential functions

$$u_{\mu_t} : \mathbb{R} \rightarrow \mathbb{R}, \quad u_{\mu_t}(x) := \int |x - y| \mu_t(dy).$$

The following properties are elementary:

- (i)  $u_{\mu_t}$  is nonnegative and convex,
- (ii)  $\partial^+ u_{\mu_t}(x) - \partial^- u_{\mu_t}(x) = 2\mu_t(\{x\})$ ,
- (iii)  $\lim_{|x| \rightarrow \infty} u_{\mu_t}(x) = \infty \mathbf{1}_{\mu_t \neq 0}$ ,
- (iv)  $\lim_{|x| \rightarrow \infty} u_{\mu_t}(x) - \mu_t(\mathbb{R})|x - \text{bary}(\mu_t)| = 0$ ,



where  $\partial^+$  and  $\partial^-$  denote the right and left derivatives, respectively, and  $\text{bary}(\mu_t) = \int x d\mu_t$  is the barycenter. We can therefore extend  $u_{\mu_t}$  continuously to  $\bar{\mathbb{R}} = [-\infty, \infty]$ . The following result of Strassen is classical; cf. [39].

**Proposition 2.1.** *Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be finite measures on  $\mathbb{R}$  with finite first moments and equal total mass. The following are equivalent:*

- (i)  $\mu_0 \leq_c \dots \leq_c \mu_n$ ,
- (ii)  $u_{\mu_0} \leq \dots \leq u_{\mu_n}$ ,
- (iii)  $\mathcal{M}(\boldsymbol{\mu}) \neq \emptyset$ ,
- (iv) *there exist stochastic kernels  $\kappa_t(x_0, \dots, x_{t-1}, dx_t)$  such that*

$$\int |x_t| \kappa_t(x_0, \dots, x_{t-1}, dx_t) < \infty \text{ and } \int x_t \kappa_t(x_0, \dots, x_{t-1}, dx_t) = x_{t-1}$$
*for all  $(x_0, \dots, x_t) \in \mathbb{R}^t$  and  $1 \leq t \leq n$ , and*

$$\mu_t = (\mu_0 \otimes \kappa_1 \otimes \dots \otimes \kappa_n) \circ (X_t)^{-1} \text{ for all } 0 \leq t \leq n.$$

A kernel  $\kappa_t$  with the first property in (iv) is called *martingale kernel*.

## 2.1 The One-Step Case

For the convenience of the reader, we summarize some results from [8] and [10] for the one-step problem ( $n = 1$ ) which will be used later on. In this section we write  $(\mu, \nu)$  instead of  $(\mu_0, \mu_1)$  for the given marginals in convex order.

**Definition 2.2.** The pair  $\mu \leq_c \nu$  is *irreducible* if the set  $I = \{u_\mu < u_\nu\}$  is connected and  $\mu(I) = \mu(\mathbb{R})$ . In this situation, let  $J$  be the union of  $I$  and any endpoints of  $I$  that are atoms of  $\nu$ ; then  $(I, J)$  is the *domain* of  $\mathcal{M}(\mu, \nu)$ .

The first result is a decomposition of the transport problem into irreducible parts; cf. [8, Theorem 8.4].

**Proposition 2.3.** *Let  $\mu \leq_c \nu$  and let  $(I_k)_{1 \leq k \leq N}$  be the (open) components of  $\{u_\mu < u_\nu\}$ , where  $N \in \{0, 1, \dots, \infty\}$ . Set  $I_0 = \mathbb{R} \setminus \cup_{k \geq 1} I_k$  and  $\mu_k = \mu|_{I_k}$  for  $k \geq 0$ , so that  $\mu = \sum_{k \geq 0} \mu_k$ . Then, there exists a unique decomposition  $\nu = \sum_{k \geq 0} \nu_k$  such that*

$$\mu_0 = \nu_0 \quad \text{and} \quad \mu_k \leq_c \nu_k \quad \text{for all } k \geq 1,$$

*and this decomposition satisfies  $I_k = \{u_{\mu_k} < u_{\nu_k}\}$  for all  $k \geq 1$ . Moreover, any  $P \in \mathcal{M}(\mu, \nu)$  admits a unique decomposition  $P = \sum_{k \geq 0} P_k$  such that  $P_k \in \mathcal{M}(\mu_k, \nu_k)$  for all  $k \geq 0$ .*

We observe that the measure  $P_0$  in Proposition 2.3 transports  $\mu_0$  to itself and is concentrated on  $\Delta_0 := \Delta \cap I_0^2$  where  $\Delta = \{(x, x) : x \in \mathbb{R}\}$  is the diagonal. Thus, the transport problem with index  $k = 0$  is not actually an irreducible one, but we shall nevertheless refer to  $(I_0, I_0)$  as the domain of this problem. When we want to emphasize the distinction, we call  $(I_0, I_0)$  the *diagonal domain* and  $(I_k, J_k)_{k \geq 1}$  the *irreducible domains* of  $\mathcal{M}(\mu, \nu)$ . Similarly, the sets  $V_k := I_k \times J_k$ ,  $k \geq 1$  will be called the *irreducible components* and  $V_0 := \Delta_0$  will be called the *diagonal component* of  $\mathcal{M}(\mu, \nu)$ . This terminology refers to the following result of [10, Theorem 3.2] which essentially states that the components are the only sets that can be charged by a martingale transport. We call a set  $B \subseteq \mathbb{R}^2$   $\mathcal{M}(\mu, \nu)$ -polar if it is  $P$ -null for all  $P \in \mathcal{M}(\mu, \nu)$ .

**Proposition 2.4.** *Let  $\mu \leq_c \nu$  and let  $B \subseteq \mathbb{R}^2$  be a Borel set. Then  $B$  is  $\mathcal{M}(\mu, \nu)$ -polar if and only if there exist a  $\mu$ -nullset  $N_\mu$  and a  $\nu$ -nullset  $N_\nu$  such that*

$$B \subseteq (N_\mu \times \mathbb{R}) \cup (\mathbb{R} \times N_\nu) \cup \left( \bigcup_{k \geq 0} V_k \right)^c.$$

The following result of [10, Lemma 3.3] will also be useful; it is the main ingredient in the proof of the preceding proposition.

**Lemma 2.5.** *Let  $\mu \leq_c \nu$  be irreducible and let  $\pi$  be a finite measure on  $\mathbb{R}^2$  whose marginals  $\pi_1, \pi_2$  satisfy<sup>1</sup>  $\pi_1 \leq \mu$  and  $\pi_2 \leq \nu$ . Then, there exists  $P \in \mathcal{M}(\mu, \nu)$  such that  $P$  dominates  $\pi$  in the sense of absolute continuity.*

### 3 The Polar Structure

The goal of this section is to identify all obstructions to martingale transports imposed by the marginals  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ , and thus, conversely, the sets that can indeed be charged. We recall that a subset  $B$  of  $\mathbb{R}^{n+1}$  is called  $\mathcal{M}(\boldsymbol{\mu})$ -polar if it is a  $P$ -nullset for all  $P \in \mathcal{M}(\boldsymbol{\mu})$ . The result for the one-step case in Proposition 2.4 already exhibits an obvious type of polar set  $B \subseteq \mathbb{R}^{n+1}$ : if for some  $t$  there is an  $\mathcal{M}(\mu_{t-1}, \mu_t)$ -polar set  $B' \subseteq \mathbb{R}^2$  such that  $B \subseteq \mathbb{R}^{t-1} \times B' \times \mathbb{R}^{n-t}$ , then  $B$  must be  $\mathcal{M}(\boldsymbol{\mu})$ -polar. The following shows that unions of such sets are in fact the only polar sets of  $\mathcal{M}(\boldsymbol{\mu})$ .

---

<sup>1</sup>By  $\pi_1 \leq \mu$  we mean that  $\pi_1(A) \leq \mu(A)$  for every Borel set  $A \subseteq \mathbb{R}$ .

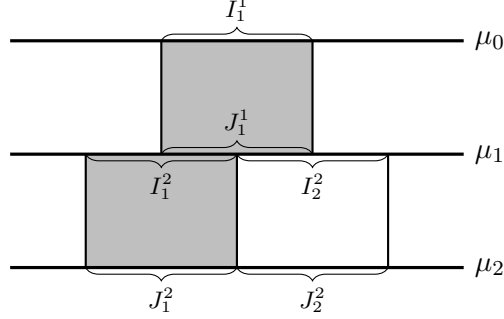


Figure 2: The shaded area represents  $V_{\mathbf{k}}$  for  $\mathbf{k} = (1, 1)$ .

**Theorem 3.1** (Polar Structure). *Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order. Then a Borel set  $B \subseteq \mathbb{R}^{n+1}$  is  $\mathcal{M}(\boldsymbol{\mu})$ -polar if and only if there exist  $\mu_t$ -nullsets  $N_t$  such that*

$$B \subseteq \bigcup_{t=0}^n (X_t)^{-1}(N_t) \cup \bigcup_{t=1}^n (X_{t-1}, X_t)^{-1} \left( \bigcup_{k \geq 0} V_k^t \right)^c \quad (3.1)$$

where  $(V_k^t)_{k \geq 1}$  are the irreducible components of  $\mathcal{M}(\mu_{t-1}, \mu_t)$  and  $V_0^t$  is the corresponding diagonal component.

Before stating the proof, we introduce some additional terminology. The second part of (3.1) can be expressed as

$$\begin{aligned} \bigcup_{t=1}^n (X_{t-1}, X_t)^{-1} \left( \bigcup_{k \geq 0} V_k^t \right)^c &= \left( \bigcap_{t=1}^n \bigcup_{k \geq 0} (X_{t-1}, X_t)^{-1}(V_k^t) \right)^c \\ &= \left( \bigcup_{k_1, \dots, k_n \geq 0} \bigcap_{t=1}^n (X_{t-1}, X_t)^{-1}(V_{k_t}^t) \right)^c. \end{aligned} \quad (3.2)$$

For every  $\mathbf{k} = (k_1, \dots, k_n)$ , the set

$$V_{\mathbf{k}} = \bigcap_{t=1}^n (X_{t-1}, X_t)^{-1}(V_{k_t}^t) \subseteq \mathbb{R}^{n+1}$$

as occurring in the last expression of (3.2) will be referred to as an *irreducible component* of  $\mathcal{M}(\boldsymbol{\mu})$ ; these sets are disjoint since  $V_k^t \cap V_{k'}^t = \emptyset$  for  $k \neq k'$ .

Moreover, we call their union

$$\mathcal{V} = \cup_{\mathbf{k}} V_{\mathbf{k}}$$

the *effective domain* of  $\mathcal{M}(\boldsymbol{\mu})$ .

Roughly speaking, an irreducible component  $V_{\mathbf{k}}$  is a chain of irreducible components from the individual steps  $(t-1, t)$ . Our terminology is somewhat abusive: in the one-step case considered in [8, 10], it was possible and useful to decompose the transport problem into its irreducible components and study those separately to a large extent; cf. Proposition 2.3. This is impossible in the multistep case, as illustrated by the following example.

**Example 3.2.** Consider the two-step martingale transport problem with marginals  $\mu_0 = \delta_0$ ,  $\mu_1 = \frac{1}{2}(\delta_{-1} + \delta_1)$  and  $\mu_2 = \frac{1}{4}(\delta_{-2} + 2\delta_0 + \delta_2)$ . Then the irreducible components are given by

$$\begin{aligned} V_{00} &= \{(x, x, x) : x \notin (-2, 2)\} \\ V_{01} &= \{(x, x) : x \in (-2, -1]\} \times [-2, 0] \\ V_{02} &= \{(x, x) : x \in [1, 2)\} \times [0, 2] \\ V_{10} &= (-1, 1) \times \{0\} \times \{0\} \\ V_{11} &= (-1, 1) \times [-1, 0] \times [-2, 0] \\ V_{12} &= (-1, 1) \times (0, 1] \times [0, 2]. \end{aligned}$$

There is only one martingale transport  $P \in \mathcal{M}(\boldsymbol{\mu})$ , given by

$$P = \frac{1}{4}(\delta_{(0,-1,-2)} + \delta_{(0,-1,0)} + \delta_{(0,1,0)} + \delta_{(0,1,2)}).$$

While  $P$  is supported on  $V_{11} \cup V_{12}$ , it cannot be decomposed into two martingale parts that are supported on  $V_{11}$  and  $V_{12}$ , respectively:  $V_{11}$  and  $V_{12}$  are disjoint, but  $P|_{V_{11}} = \frac{1}{4}(\delta_{(0,-1,-2)} + \delta_{(0,-1,0)})$  is not a martingale.

The main step in the proof of Theorem 3.1 will be the following lemma.

**Lemma 3.3.** *Let  $V_{\mathbf{k}}$  be an irreducible component of  $\mathcal{M}(\boldsymbol{\mu})$  and consider a measure  $\pi$  concentrated on  $V_{\mathbf{k}}$  such that  $\pi_t \leq \mu_t$  for  $t = 0, \dots, n$ . Then there exists a transport  $P \in \mathcal{M}(\boldsymbol{\mu})$  which dominates  $\pi$  in the sense of absolute continuity.*

Deferring the proof, we first show how this implies the theorem.

*Proof of Theorem 3.1.* Clearly  $(X_t)^{-1}(N_t)$  is  $\mathcal{M}(\boldsymbol{\mu})$ -polar for  $t = 0, \dots, n$  and  $(X_{t-1}, X_t)^{-1} \left( \bigcup_{k \geq 0} V_k^t \right)^c$  is  $\mathcal{M}(\boldsymbol{\mu})$ -polar for  $t = 1, \dots, n$ . This shows that (3.1) is sufficient for  $B \subseteq \mathbb{R}^{n+1}$  to be  $\mathcal{M}(\boldsymbol{\mu})$ -polar.

Conversely, suppose that (3.1) does not hold; we show that  $B$  is not  $\mathcal{M}(\boldsymbol{\mu})$ -polar. In view of (3.2), we may assume that

$$B \subseteq V_{\mathbf{k}} = \bigcap_{t=1}^n (X_{t-1}, X_t)^{-1}(V_{k_t}^t)$$

for some  $\mathbf{k}$  by passing to a subset of  $B$  if necessary. Passing to a further subset, we may also assume that there are no  $\mu_t$ -nullsets  $N_t$  such that  $B \subseteq \bigcup_{t=0}^n (X_t)^{-1}(N_t)$ . By a result of classical optimal transport [4, Proposition 2.1], this entails that  $B$  is not  $\Pi(\boldsymbol{\mu})$ -polar; i.e. we can find a measure  $\rho \in \Pi(\boldsymbol{\mu})$  such that  $\rho(B) > 0$ . But then  $\pi := \rho|_{V_{\mathbf{k}}}$  satisfies the assumptions of Lemma 3.3 which yields  $P \in \mathcal{M}(\boldsymbol{\mu})$  such that  $P \gg \pi$ . In particular,  $P(B) > 0$  and  $B$  is not  $\mathcal{M}(\boldsymbol{\mu})$ -polar.  $\square$

### 3.1 Proof of Lemma 3.3

The reasoning for Lemma 3.3 follows an induction on the number  $n$  of time steps; its rigorous formulation requires a certain amount of control over subsequent steps of the transport problem. Thus, we first state a more quantitative version of (the core part of) the lemma that is tailored to the inductive argument.

**Lemma 3.4.** *Let  $\pi$  be a measure on  $\mathbb{R}^{n+1}$  with  $\pi_t \leq \mu_t$  for  $t = 0, \dots, n$  which is concentrated on the effective domain  $\mathcal{V}$  of  $\mathcal{M}(\boldsymbol{\mu})$ . Moreover, let  $(\{O_k\}_{k \geq 1}, \{L_k\}_{k \geq 1})$  be an at most countable partition of  $\mathbb{R}$  consisting of intervals<sup>2</sup> with the property that  $\mu_n(O_k) > 0$  and  $O_k$  is open for all  $k \geq 1$ .*

*Assume that the marginals  $\pi_t$  of  $\pi$  satisfy the following condition: for every  $0 \leq t \leq n$  there exist finitely many compact intervals  $K_{t,1}, \dots, K_{t,N_t}$  such that  $\pi_t$  is concentrated on  $\bigcup_i K_{t,i}$  and moreover,*

- *if  $t < n$ : each  $K_{t,i}$  satisfies one of the following,*

(A1)  $K_{t,i} \subseteq I$  for some irreducible domain  $(I, J)$  of  $\mathcal{M}(\mu_t, \mu_{t+1})$  that is not the diagonal domain,

(B1)  $K_{t,i} \subseteq I_0$ , the domain of the diagonal component of  $\mathcal{M}(\mu_t, \mu_{t+1})$ ;

---

<sup>2</sup>An interval is defined as a connected subset of  $\mathbb{R}$ .

- if  $t = n$ : each  $K_{t,i}$  satisfies one of the following,

(A2)  $K_{t,i} \subseteq O_k$  for some  $k \geq 1$ ,

(B2)  $K_{t,i} \subseteq L_k$  for some  $k \geq 1$ .

Then, there is a martingale measure  $Q$  on  $\mathbb{R}^{n+1}$  that dominates  $\pi$  and has marginals  $Q_t \leq \mu_t$  for  $0 \leq t \leq n$  which again satisfy the above condition. Moreover,  $Q$  can be chosen such that  $dQ = g d\pi + d\sigma$  where the density  $g$  is bounded and the measure  $\sigma$  is singular with respect to  $\pi$ .

*Proof.* We proceed by induction on the number of marginals which we assume to be probability measures. For  $n = 0$  we set  $Q = \pi$  to obtain a 0-step martingale.

Now, consider  $n \geq 1$  and assume the lemma has already been shown for  $(n - 1)$ -step measures. We disintegrate

$$\pi = \pi' \otimes \kappa(x_0, \dots, x_{n-1}, dx_n) \quad (3.3)$$

and observe that  $\pi'$  satisfies the conditions of the lemma with the countable families

$$\{\tilde{O}_k\}_{k \geq 1} = \{I_k\}_{k \geq 1}$$

$$\{\tilde{L}_k\}_{k \geq 1} = \{\text{connected components of } I_0 \cap O_l \text{ and } I_0 \cap L_l \text{ for some } l \geq 1\}$$

where  $\{I_k\}_{k \geq 1}$  are the open intervals from the irreducible domains  $(I_k, J_k)$  of  $\mathcal{M}(\mu_{n-1}, \mu_n)$  and  $I_0$  denotes the corresponding diagonal domain. (To see this, we use that  $\pi$  must be an identical transport on  $I_0$  by the definition of  $\mathcal{V}$ .) By the induction assumption, we then find a martingale measure  $Q' \gg \pi'$  on  $\mathbb{R}^n$  with the stated properties. In particular, the marginal  $Q'_{n-1}$  is supported by finitely many compact intervals  $K$  satisfying (A2) or (B2); that is, either  $K \subseteq I$  for some irreducible domain  $(I, J)$  of  $\mathcal{M}(\mu_{n-1}, \mu_n)$ , or  $K \subseteq I_0 \cap O_k$  or  $K \subseteq I_0 \cap L_k$ , for the diagonal domain  $I_0$  of  $\mathcal{M}(\mu_{n-1}, \mu_n)$ .

We shall construct a martingale kernel  $\hat{\kappa}$  by suitably manipulating  $\kappa$ . Let us observe that since  $\pi$  is concentrated on  $\mathcal{V}$ , the following hold for  $\pi'$ -a.e.  $\mathbf{x} = (x_0, \dots, x_{n-1}) \in \mathbb{R}^n$ :

- $\kappa(\mathbf{x}, \cdot) = \delta_{x_{n-1}}$  whenever  $x_{n-1} \in I_0$ ,
- $\kappa(\mathbf{x}, \cdot)$  is concentrated on intervals  $K_{n,i}$  as in the statement of the lemma with  $K_{n,i} \subseteq J_k$  for  $x_{n-1} \in I_k$  with  $k \geq 1$  and  $Q'_{n-1}(I_k) > 0$ .

By changing  $\kappa$  on a  $\pi'$ -nullset, we may assume that these two properties hold for all  $\mathbf{x} \in \mathbb{R}^n$ .

*Step 1.* Next, we argue that we may change  $Q'$  and  $\kappa$  such that the marginal  $(Q' \otimes \kappa)_n = (Q' \otimes \kappa) \circ X_n^{-1}$  satisfies

$$(Q' \otimes \kappa)_n \leq \mu_n. \quad (3.4)$$

Indeed, recall that  $dQ' = dQ'_{abs} + d\sigma' = g'd\pi' + d\sigma'$  where the density  $g'$  is bounded and  $\sigma'$  is singular with respect to  $\pi'$ . Using the Lebesgue decomposition theorem, we find a Borel set  $A \subseteq \mathbb{R}^n$  such that  $\sigma'(A) = \sigma'(\mathbb{R}^n)$  and  $\pi'(A) = 0$ .

By scaling  $Q'$  with a constant we may assume that  $g' \leq 1/2$ . As  $\pi_n \leq \mu_n$ , the marginal  $(Q'_{abs} \otimes \kappa)_n$  is then bounded by  $\frac{1}{2}\mu_n$ , and it remains to bound  $(\sigma' \otimes \kappa)_n$  in the same way.

Note that  $Q'_{n-1} \leq \mu_{n-1}$  implies  $\sigma'_{n-1} \leq \mu_{n-1}$ . We may change  $\kappa$  arbitrarily on the set  $A$  without invalidating (3.3). Indeed, for each irreducible component  $(I_k, J_k)$  of  $\mathcal{M}(\mu_{n-1}, \mu_n)$  we choose and fix a compact interval  $K_k \subseteq J_k$  with  $\mu_n(K_k) > 0$  which is contained in one of the sets  $O_l$  or  $L_l$ ; this is possible since the latter form a countable partition of  $\mathbb{R}$ .

For  $\mathbf{x} = (x_0, \dots, x_{n-1}) \in A$  such that  $x_{n-1} \in I_k$  we then define

$$\kappa(\mathbf{x}, \cdot) := \frac{1}{\mu_n(K_k)} \mu_n|_{K_k}.$$

Set  $\epsilon_k = \mu_n(K_k)/\mu_{n-1}(I_k)$ . Then

$$\epsilon := \inf_{k: Q'_{n-1}(I_k) > 0} \epsilon_k \wedge 1$$

is strictly positive because there are only finitely many  $k$  with  $Q'_{n-1}(I_k) > 0$  (this is the purpose of the elaborate induction assumption). As  $\sigma'_{n-1} \leq \mu_{n-1}$ , we may scale  $Q'$  once again to obtain  $\sigma'_{n-1} \leq \frac{\epsilon}{6}\mu_{n-1}$ . We now have

$$(\sigma'|_{\mathbb{R}^{n-1} \times I_k} \otimes \kappa)_n = \sigma'_{n-1}(I_k) \frac{1}{\mu_n(K_k)} \mu_n|_{K_k} \leq \frac{\epsilon}{6} \frac{\mu_{n-1}(I_k)}{\mu_n(K_k)} \mu_n|_{K_k} \leq \frac{1}{6} \mu_n|_{K_k}.$$

For the diagonal domain  $I_0$  the corresponding inequality holds because we have  $\kappa(\mathbf{x}, \cdot) = \delta_{x_{n-1}}$  for  $x_{n-1} \in I_0$  and  $\sigma'_{n-1}|_{I_0} \leq \frac{1}{6}\mu_{n-1}|_{I_0} = \frac{1}{6}\mu_n|_{I_0}$ . As a consequence, we have  $(\sigma' \otimes \kappa)_n \leq \frac{1}{2}\mu_n$  as desired, so that we may assume (3.4) in what follows.

*Step 2.* We now construct a martingale kernel  $\hat{\kappa}$  such that  $Q = Q' \otimes \hat{\kappa}$  has the required properties. For a fixed irreducible component  $(I_k, J_k)$  we have

that  $Q'_{n-1}|_{I_k} = Q'_{n-1}|_K$  for some compact  $K \subseteq I_k$ . We can find compact intervals  $B^-, B^+ \subseteq J_k$  with  $\mu_n(B^-) > 0$  and  $\mu_n(B^+) > 0$  such that  $B^-$  is to the left of  $K$  and  $B^+$  is to the right of  $K$ , in the sense that  $x < y < z$  for  $x \in B^-$ ,  $y \in K$  and  $z \in B^+$ . By replacing  $B^\pm$  with a compact subinterval of  $B^\pm \cap O_k$  or  $B^\pm \cap L_k$ , we may assume that each of the sets  $B^+, B^-$  is contained in one of the intervals  $O_k$  or  $L_k$ .

Next, we define two nonnegative functions  $\mathbf{x} \mapsto \varepsilon^-(\mathbf{x}), \varepsilon^+(\mathbf{x})$  for  $\mathbf{x} = (x_0, \dots, x_{n-1}) \in \mathbb{R}^{n-1} \times K$  as follows:

- for  $\mathbf{x}$  such that  $\text{bary}(\kappa(\mathbf{x}, \cdot)) < x_{n-1}$ , let  $\varepsilon^+$  be the unique number such that  $\kappa(\mathbf{x}, \cdot) + \varepsilon^+(\mathbf{x}) \cdot \mu_n|_{B^+}$  has barycenter  $x_{n-1}$ ,
- for  $\mathbf{x}$  such that  $\text{bary}(\kappa(\mathbf{x}, \cdot)) > x_{n-1}$ , let  $\varepsilon^-$  be the unique number such that  $\kappa(\mathbf{x}, \cdot) + \varepsilon^-(\mathbf{x}) \cdot \mu_n|_{B^-}$  has barycenter  $x_{n-1}$ ,
- $\varepsilon^\pm(\mathbf{x}) = 0$  otherwise.

Observe that these numbers always exist because  $B^-$  and  $B^+$  have positive mass and positive distance from the points  $x_{n-1} \in K$ . We now define the martingale kernel  $\hat{\kappa}$  by

$$\hat{\kappa}(\mathbf{x}) := c(\varepsilon^- \cdot \mu_n|_{B^-} + \kappa + \varepsilon^+ \cdot \mu_n|_{B^+})$$

where  $0 < c < 1$  is a normalizing constant such that  $\hat{\kappa}$  is again a stochastic kernel. We also define  $\hat{\kappa}(\mathbf{x}) = \kappa(\mathbf{x})$  for  $\mathbf{x}$  on the diagonal domain.

For each  $k \geq 1$ , let  $B_k^\pm$  denote the sets associated with  $I_k$  as above. Once again, the number

$$C := \inf_{k: Q'_{n-1}(I_k) > 0} [\mu_n(B_k^-) \wedge \mu_n(B_k^+)]$$

is strictly positive because there are only finitely many  $k$  with  $Q'_{n-1}(I_k) > 0$ . We can now define

$$Q := C \cdot (Q' \otimes \hat{\kappa}).$$

Then  $Q$  is a martingale transport whose marginals satisfy  $Q_t \leq Q'_t \leq \mu_t$  for  $0 \leq t \leq n-1$  whereas  $Q_n \leq \mu_n$  by (3.4), the construction of  $\hat{\kappa}$  and the choice of  $C$ . Moreover,  $Q_n$  is concentrated on the support of  $\kappa$  and the sets  $B_k^\pm$  for the finitely many  $k$  such that  $Q'_{n-1}(I_k) > 0$ . In particular,  $Q_n$  is concentrated on a finite family of compact subintervals of  $\{O_k\}_{k \geq 1}$  and  $\{L_k\}_{k \geq 1}$ .

It remains to check that  $Q$  has the required decomposition. Indeed, the definition of  $\hat{\kappa}$  can be written as

$$\hat{\kappa} = c\kappa + (1-c)\kappa^\perp$$



where  $\kappa^\perp$  is singular to  $\kappa$ . Recalling the decomposition  $Q' = Q'_{abs} + \sigma'$ , we then have

$$Q' \otimes \hat{\kappa} = cQ'_{abs} \otimes \kappa + (1 - c)Q'_{abs} \otimes \kappa^\perp + \sigma' \otimes \hat{\kappa}.$$

The last two terms are singular with respect to  $\pi = \pi' \otimes \kappa$ , and the first term is absolutely continuous with bounded density.  $\square$

*Proof of Lemma 3.3.* Let  $\pi$  be a measure with marginals  $\pi_t \leq \mu_t$  for all  $t$  which is concentrated on some irreducible component  $V_{\mathbf{k}}$  and thus, in particular, on the effective domain  $\mathcal{V}$ .

We first decompose  $\pi = \sum_{m=1}^{\infty} 2^{-m} \pi^m$  such that each  $\pi^m$  satisfies the requirements of Lemma 3.4 for the partition  $(\{O_k\}_{k \geq 1}, \{L_k\}_{k \geq 1})$  consisting of the single element  $O_1 = \mathbb{R}$ . This can be done by defining  $\pi^m$  as the restriction of  $\pi$  to a suitable subset of  $\mathbb{R}^{n+1}$ . For each of the measures  $\pi^m$ , Lemma 3.4 yields a martingale measure  $Q^m \gg \pi^m$ . We show below that there exist  $P^m \in \mathcal{M}(\boldsymbol{\mu})$  such that  $P^m \gg Q^m$ , and then  $P := \sum 2^{-m} P^m$  satisfies  $P \in \mathcal{M}(\boldsymbol{\mu})$  and  $P \gg \pi$  as desired.

To complete the proof, it remains to show that for fixed  $m \geq 1$  there exist  $0 < \epsilon_m < 1$  and  $\bar{Q}^m \in \mathcal{M}(\boldsymbol{\mu} - \epsilon_m(Q_0^m, \dots, Q_n^m))$ , as we may then conclude by setting  $P^m := \epsilon_m Q^m + \bar{Q}^m \in \mathcal{M}(\boldsymbol{\mu})$ . By Proposition 2.1, the set  $\mathcal{M}(\boldsymbol{\mu} - \epsilon_m(Q_0^m, \dots, Q_n^m))$  is nonempty if the marginals are in convex order, or equivalently if the potential functions satisfy

$$u_{\mu_{t-1}} - \epsilon_m u_{Q_{t-1}^m} \leq u_{\mu_t} - \epsilon_m u_{Q_t^m}$$

for  $t = 1, \dots, n$ . Thus, it suffices to find  $\epsilon_m > 0$  with this property for fixed  $t$ , and we have reduced to a question about a one-step martingale transport problem. Indeed, the existence of such  $\epsilon_m > 0$  follows from the fact that the continuous functions  $u_{\mu_{t-1}}$  and  $u_{\mu_t}$  are uniformly separated on compact subsets of  $\{u_{\mu_{t-1}} < u_{\mu_t}\}$ ; we refer to the proof of [10, Lemma 3.3] for the details of the argument.  $\square$

## 4 The Dual Space

In this section we introduce the domain of the dual optimization problem and show that it has a certain closedness property. The latter will be crucial for the duality theorem in the subsequent section.

We shall need a generalized notion of integrability for the elements of the dual space. To this end, we first recall the integral for concave functions as detailed in [10, Section 4.1].

**Definition 4.1.** Let  $\mu \leq_c \nu$  be irreducible with domain  $(I, J)$  and let  $\chi : J \rightarrow \mathbb{R}$  be a concave function. We define

$$(\mu - \nu)(\chi) := \frac{1}{2} \int_I (u_\mu - u_\nu) d\chi'' + \int_{J \setminus I} |\Delta\chi| d\nu \in [0, \infty]$$

where  $-\chi''$  is the (locally finite) second derivative measure of  $-\chi$  on  $I$  and  $|\Delta\chi|$  is the absolute magnitude of the jumps of  $\chi$  at the boundary points  $J \setminus I$ .

**Remark 4.2.** As shown in [10, Lemma 4.1], this integral is well-defined and satisfies

$$(\mu - \nu)(\chi) = \int_I \left[ \chi(x) - \int_J \chi(y) \kappa(x, dy) \right] \mu(dx)$$

for any  $P = \mu \otimes \kappa \in \mathcal{M}(\mu, \nu)$ . Moreover, it coincides with the difference  $\mu(\chi) - \nu(\chi)$  of the usual integrals when  $\chi \in L^1(\mu) \cap L^1(\nu)$ .

For later reference, we record two more properties of the integral.

**Lemma 4.3.** Let  $\mu \leq_c \nu$  be irreducible with domain  $(I, J)$  and let  $\chi : J \rightarrow \mathbb{R}$  be concave.

- (i) Assume that  $I$  has a finite right endpoint  $r$  and  $\chi(a) = \chi'(a) = 0$  for some  $a \in I$ . Then  $\chi \leq 0$  and  $\chi \mathbf{1}_{[a, \infty)}$  is concave. If  $\nu$  has an atom at  $r$ , then

$$\chi(r) \geq -\frac{C}{\nu(\{r\})} (\mu - \nu)(\chi \mathbf{1}_{[a, \infty)})$$

for a constant  $C \geq 0$  depending only on  $\mu, \nu$ .

- (ii) For  $a, b \in \mathbb{R}$ , the concave function  $\bar{\chi}(x) := \chi(x) + ax + b$  satisfies

$$(\mu - \nu)(\bar{\chi}) = (\mu - \nu)(\chi).$$

*Proof.* The first part is [10, Remark 4.6] and the second part follows directly from  $\bar{\chi}'' = \chi''$  and  $\Delta\bar{\chi} = \Delta\chi$ .  $\square$

Let us now return to the multistep case with a vector  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  of measures in convex order and introduce  $\boldsymbol{\mu}(\boldsymbol{\phi}) := \sum_{t=0}^n \mu_t(\phi_t)$  in cases where we do not necessarily have  $\phi_t \in L^1(\mu_t)$ . As mentioned previously, the multistep transport problem does not decompose into irreducible components, forcing us to directly give a global definition of the integral.

**Definition 4.4.** Let  $\boldsymbol{\phi} = (\phi_0, \dots, \phi_n)$  be a vector of Borel functions  $\phi_t : \mathbb{R} \rightarrow \bar{\mathbb{R}}$ . A vector  $\boldsymbol{\chi} = (\chi_1, \dots, \chi_n)$  of Borel functions  $\chi_t : \mathbb{R} \rightarrow \mathbb{R}$  is called a *concave moderator* for  $\boldsymbol{\phi}$  if for  $1 \leq t \leq n$ ,

- (i)  $\chi_t|_J$  is concave for every domain  $(I, J)$  of an irreducible component of  $\mathcal{M}(\mu_{t-1}, \mu_t)$ ,
- (ii)  $\chi_t|_{I_0} \equiv 0$  for the diagonal domain  $I_0$  of  $\mathcal{M}(\mu_{t-1}, \mu_t)$ ,
- (iii)  $\phi_t - \chi_{t+1} + \chi_t \in L^1(\mu_t)$ ,

where  $\chi_{n+1} \equiv 0$ . We also convene that  $\chi_0 \equiv 0$ . The *moderated integral* of  $\phi$  is then defined by

$$\mu(\phi) := \sum_{t=0}^n \mu_t(\phi_t - \chi_{t+1} + \chi_t) + \sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t) \in (-\infty, \infty], \quad (4.1)$$

where  $(\mu_{t-1} - \mu_t)^k(\chi_t)$  denotes the integral of Definition 4.1 on the  $k$ -th irreducible component of  $\mathcal{M}(\mu_{t-1}, \mu_t)$ .

**Remark 4.5.** The moderated integral is independent of the choice of the moderator  $\chi$ . To see this, consider a second moderator  $\chi'$  for  $\phi$ ; then we have  $(\chi'_{t+1} - \chi_{t+1}) - (\chi'_t - \chi_t) \in L^1(\mu_t)$ . Using Remark 4.2 with arbitrary  $\kappa_t$  such that  $\mu_{t-1} \otimes \kappa_t \in \mathcal{M}(\mu_{t-1}, \mu_t)$  for  $1 \leq t \leq n$ , as well as Fubini's theorem for kernels,

$$\begin{aligned} & \sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t) - (\mu_{t-1} - \mu_t)^k(\chi'_t) \\ &= \int \cdots \int \sum_{t=1}^n \chi_t(x_{t-1}) - \chi_t(x_t) \kappa_n(x_{n-1}, dx_n) \cdots \kappa_1(x_0, dx_1) \mu_0(dx_0) \\ & \quad - \int \cdots \int \sum_{t=1}^n \chi'_t(x_{t-1}) - \chi'_t(x_t) \kappa_n(x_{n-1}, dx_n) \cdots \kappa_1(x_0, dx_1) \mu_0(dx_0) \\ &= \sum_{t=0}^n \mu_t((\chi'_{t+1} - \chi_{t+1}) - (\chi'_t - \chi_t)). \end{aligned}$$

It now follows that (4.1) yields the same value for both moderators.

For later reference, we also record the following property.

**Remark 4.6.** If  $\chi$  is a concave moderator, Definition 4.4 (ii) implies that

$$\chi_t = \sum_{k \geq 1} \chi_t|_{I_k^t} = \sum_{k \geq 1} \chi_t|_{J_k^t}$$

where  $(I_k^t, J_k^t)$  is the  $k$ -th irreducible domain of  $\mathcal{M}(\mu_{t-1}, \mu_t)$ .

Next, we introduce the space of functions which have a finite integral in the moderated sense.

**Definition 4.7.** We denote by  $L^c(\boldsymbol{\mu})$  the space of all vectors  $\phi$  admitting a concave moderator  $\chi$  with  $\sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k (\chi_t) < \infty$ .

It follows that  $\boldsymbol{\mu}(\phi)$  is finite for  $\phi \in L^c(\boldsymbol{\mu})$ , and we have  $\boldsymbol{\mu}(\phi) = \sum_t \mu_t(\phi_t)$  for  $\phi \in \Pi_{t=0}^n L^1(\mu_t)$ . The definition is also consistent with the expectation under martingale transports, in the following sense.

**Lemma 4.8.** Let  $\phi \in L^c(\boldsymbol{\mu})$  and let  $H = (H_1, \dots, H_n)$  be  $\mathbb{F}$ -predictable. If

$$\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n$$

is bounded from below on the effective domain  $\mathcal{V}$  of  $\mathcal{M}(\boldsymbol{\mu})$ , then

$$\boldsymbol{\mu}(\phi) = P \left[ \sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \right], \quad P \in \mathcal{M}(\boldsymbol{\mu}).$$

*Proof.* Let  $P \in \mathcal{M}(\boldsymbol{\mu})$ , let  $\chi$  be a concave moderator for  $\phi$ , and assume without loss of generality that 0 is the lower bound. Using Remark 4.6, we have that  $\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n$  equals

$$\sum_{t=0}^n (\phi_t - \chi_{t+1} + \chi_t)(X_t) + \sum_{t=1}^n \sum_{k \geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_n \geq 0.$$

By assumption, the functions  $(\phi_t - \chi_{t+1} + \chi_t)(X_t)$  are  $P$ -integrable. Therefore, the negative part of the remaining expression must also be  $P$ -integrable. Writing  $P_t := P \circ (X_0, \dots, X_t)^{-1}$  and using that  $(\chi_t|_{J_k^t})^+$  has linear growth, we see that for any disintegration  $P = P_{n-1} \otimes \kappa_n$ ,

$$\begin{aligned} & \int \sum_{t=1}^n \sum_{k \geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_n \kappa_n(X_0, \dots, X_{n-1}, dX_n) \\ &= \sum_{t=1}^{n-1} \sum_{k \geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_{n-1} \\ & \quad + \sum_{k \geq 1} \int [\chi_n|_{I_k^n}(X_{n-1}) - \chi_n|_{J_k^n}(X_n)] \kappa_n(X_0, \dots, X_{n-1}, dX_n). \end{aligned}$$

Iteratively integrating with kernels such that  $P_t = P_{t-1} \otimes \kappa_t$  and observing that we can apply Fubini's theorem to  $\sum_{t=1}^n \sum_{k \geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_n$  as its negative part is  $P$ -integrable, we obtain

$$P \left[ \sum_{t=1}^n \sum_{k \geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_n \right] = \sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t)$$

and the result follows.  $\square$

We can now define our dual space. It will be convenient to work with non-negative reward functions  $f$  for the moment—we shall relax this constraint later on; cf. Remark 5.3.

**Definition 4.9.** Let  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$ . We denote by  $\mathcal{D}_\mu(f)$  the set of all pairs  $(\phi, H)$  where  $\phi \in L^c(\mu)$  and  $H = (H_1, \dots, H_n)$  is an  $\mathbb{F}$ -predictable process such that

$$\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \geq f \quad \text{on } \mathcal{V}.$$

By Lemma 4.8, the expectation of the left hand side under any  $P \in \mathcal{M}(\mu)$  is given by the moderated integral  $\mu(\phi)$ ; this will be seen as the dual cost of  $(\phi, H)$  when we consider the dual problem  $\inf_{(\phi, H) \in \mathcal{D}_\mu(f)} \mu(\phi)$  in Section 5 below.

The following closedness property is the key result about the dual space.

**Proposition 4.10.** *Let  $f^m : \mathbb{R}^{n+1} \rightarrow [0, \infty]$ ,  $m \geq 1$  be a sequence of functions such that*

$$f^m \rightarrow f \quad \text{pointwise}$$

*and let  $(\phi^m, H^m) \in \mathcal{D}_\mu(f^m)$  be such that  $\sup_m \mu(\phi^m) < \infty$ . Then there exist  $(\phi, H) \in \mathcal{D}_\mu(f)$  with*

$$\mu(\phi) \leq \liminf_{m \rightarrow \infty} \mu(\phi^m).$$

#### 4.1 Proof of Proposition 4.10

An attempt to prove Proposition 4.10 directly along the lines of [10] runs into a technical issue in controlling the concave moderators. Roughly speaking, they do not allow sufficiently many normalizations; this is related to the aforementioned fact that the multistep problem cannot be decomposed

into its components. We shall introduce a generalized dual space with families of functions indexed by the components, and prove a “lifted” version of Proposition 4.10 in this larger space. Once that is achieved, we can infer the closedness result in the original space as well. (The reader willing to admit Proposition 4.10 may skip this subsection without loss of continuity.)

**Definition 4.11.** Let  $\phi = \{\phi_t^k : 0 \leq t \leq n, k \geq 0\}$  be a family of Borel functions, consisting of one function  $\phi_t^k : J_k^t \rightarrow \mathbb{R}$  for each irreducible component  $(I_k^t, J_k^t)$  of  $\mathcal{M}(\mu_{t-1}, \mu_t)$  as indexed by  $k \geq 1$  and  $1 \leq t \leq n$ , functions  $\phi_t^0 : I_0^t \rightarrow \mathbb{R}$  for the diagonal components  $I_0^t$  indexed by  $1 \leq t \leq n$ , and a single function  $\phi_0^0 : \mathbb{R} \rightarrow \mathbb{R}$  for  $t = 0$ . Similarly, let  $\chi = \{\chi_t^k : 1 \leq t \leq n, k \geq 0\}$  be a family of functions, consisting of one concave function  $\chi_t^k : J_k^t \rightarrow \mathbb{R}$  for each irreducible component  $(I_k^t, J_k^t)$  and Borel functions  $\chi_t^0 : I_0^t \rightarrow \mathbb{R}$  for the diagonal components. We also convene that  $\chi_0^0 \equiv 0$  and define the functions  $\chi_t := \sum_{k \geq 0} \chi_t^k|_{I_k^t}$  for  $t = 1, \dots, n$ , as well as  $\chi_{n+1} \equiv 0$ .

We call  $\chi$  a *concave moderator* for  $\phi$  if

$$\phi_t^k + \chi_t^k - \chi_{t+1} \in L^1(\mu_t^k)$$

for all  $t = 0, \dots, n$  and  $k \geq 0$ , where  $\mu_t^k$  is the second marginal of the  $k$ -th irreducible component in the decomposition of  $\mathcal{M}(\mu_{t-1}, \mu_t)$  as in Proposition 2.3, and  $\mu_0^0 \equiv \mu_0$ . The generalized moderated integral is then defined by

$$\mu(\phi) := \sum_{t=0}^n \sum_{k \geq 0} \mu_t^k(\phi_t^k + \chi_t^k - \chi_{t+1}) + \sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t^k).$$

Finally, we denote by  $L^{c,g}(\mu)$  the set of all families  $\phi$  which admit a concave moderator  $\chi$  such that

$$\sum_{t=0}^n \sum_{k \geq 0} |\mu_t^k(\phi_t^k + \chi_t^k - \chi_{t+1})| + \sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t^k) < \infty.$$

We can now introduce the generalized dual space.

**Definition 4.12.** Let  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$ . We denote by  $\mathcal{D}_\mu^g(f)$  the set of all pairs  $(\phi, H)$  where  $\phi \in L^{c,g}(\mu)$ ,  $H = (H_1, \dots, H_n)$  is  $\mathbb{F}$ -predictable, and

$$\sum_{t=0}^n \phi_t^{k_t}(x_t) + (H \cdot \mathbf{x})_n \geq f(\mathbf{x})$$

for all  $\mathbf{x} = (x_0, \dots, x_n)$  and  $\mathbf{k} = (k_0, \dots, k_n)$  such that  $(x_{t-1}, x_t) \in (I_{k_t}^t, J_{k_t}^t)$  for some (irreducible or diagonal) component<sup>3</sup> and  $t = 1, \dots, n$ .

We observe that for any  $\mathbf{x} \in \mathcal{V}$  the corresponding  $\mathbf{k} = (k_0, \dots, k_n)$  is uniquely defined, where the index  $k_0 \equiv 0$  exists purely for notational convenience.

**Remark 4.13.** There are certain degrees of freedom in choosing elements of  $\mathcal{D}_{\boldsymbol{\mu}}^g(f)$  that will be important later on:

- (i) Given  $1 \leq t \leq n$ , a domain  $(I_k^t, J_k^t)$  of an irreducible component of  $\mathcal{M}(\mu_{t-1}, \mu_t)$ , and constants  $c_1, c_2 \in \mathbb{R}$ , we have that  $(\phi, H)$  is in  $\mathcal{D}_{\boldsymbol{\mu}}^g(f)$  if and only if  $(\tilde{\phi}, \tilde{H})$  is in  $\mathcal{D}_{\boldsymbol{\mu}}^g(f)$ , where  $(\tilde{\phi}, \tilde{H})$  is defined by

$$\begin{aligned}\tilde{\phi}_t^k(y) &= \phi_t^k(y) - (c_1 y - c_2) \\ \tilde{\phi}_{t-1}^{k'}(x) &= \phi_{t-1}^{k'}(x) + (c_1 x - c_2)|_{I_k^t} \\ \tilde{\phi}_s^{k'} &= \phi_s^{k'} \quad \text{for } s \notin \{t-1, t\} \\ \tilde{H}_t &= H_t + c_1|_{X_t^{-1}(I_k^t)} \\ \tilde{H}_s &= H_s \quad \text{for } s \neq t\end{aligned}$$

where  $k'$  runs over all components of the preceding step.

- (ii) We may (and will) assume without loss of generality that  $\chi_t^0 \equiv 0$  for  $t = 1, \dots, n$ . Indeed, we can pass from  $\phi$  to  $\tilde{\phi}$  defined with

$$\tilde{\phi}_t^0 := \phi_t^0 + \chi_t^0, \quad \tilde{\phi}_{t-1}^k := \phi_{t-1}^k - \chi_t^0|_{J_k^{t-1}}.$$

- (iii) Any  $(\phi, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$  induces an element  $(\phi^g, H) \in \mathcal{D}_{\boldsymbol{\mu}}^g(f)$  with  $\boldsymbol{\mu}(\phi^g) = \boldsymbol{\mu}(\phi)$  by choosing some concave moderator  $\chi$  for  $\phi$  and setting

$$\phi_t^k := \phi_t|_{J_k^t}, \quad \chi_t^k := \chi_t|_{J_k^t}.$$

Next, we establish that lifting from  $\mathcal{D}_{\boldsymbol{\mu}}(f)$  to  $\mathcal{D}_{\boldsymbol{\mu}}^g(f)$  does not change the range of dual costs.

**Proposition 4.14.** *Let  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$ . We have*

$$\{\boldsymbol{\mu}(\phi^g) : (\phi^g, H) \in \mathcal{D}_{\boldsymbol{\mu}}^g(f)\} = \{\boldsymbol{\mu}(\phi) : (\phi, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)\}.$$

---

<sup>3</sup> Given an irreducible component  $(I, J)$ , the notation  $(x, y) \in (I, J)$  means that  $(x, y) \in I \times J$ , where as for a diagonal component  $(I_0, I_0)$  it is to be understood as  $(x, y) \in (I_0 \times I_0) \cap \Delta$ .

*Proof.* Remark 4.13 (iii) shows the inclusion “ $\supseteq$ .” To show the reverse, we may apply Remark 4.13 (i) inductively (backwards in time) to modify a given pair  $(\phi^g, H) \in \mathcal{D}_\mu^g(f)$  such that  $\phi_t^k(x) = 0$  for  $x \in J_k^t \setminus I_k^t$ , for all irreducible domains  $(I_k^t, J_k^t)$  of  $\mathcal{M}(\mu_{t-1}, \mu_t)$  and  $1 \leq t \leq n$ . Here we have used that  $x \in J_k^t \setminus I_k^t$  implies  $\mu_k^t(\{x\}) > 0$ , cf. Definition 2.2, and therefore  $\phi^g \in L^{c,g}(\mu)$  implies  $\phi_t^k(x) \in \mathbb{R}$ ; that is, such endpoints can indeed be shifted to 0 by adding affine functions to  $\phi_t^k$ .

Let  $\chi^g$  be a concave moderator for  $\phi^g$ . Using Lemma 4.3 (ii) we can modify  $\chi_t^k$  to satisfy  $\chi_t^k(x) = 0$  for  $x \in J_k^t \setminus I_k^t$ , for all irreducible domains  $(I_k^t, J_k^t)$  of  $\mathcal{M}(\mu_{t-1}, \mu_t)$  and  $1 \leq t \leq n$ . Here, the finiteness of  $\chi_t^k$  at the endpoints follows from Lemma 4.3 (i) and  $(\mu_{t-1} - \mu_t)^k(\chi_t^k) < \infty$ .

Still denoting the modified dual element by  $(\phi^g, H)$ , we define  $\phi \in L^c(\mu)$  and a corresponding concave moderator  $\chi$  by

$$\phi_t(x) := \phi_t^k(x), \quad \chi_t(x) := \chi_t^k(x), \quad \text{for } x \in J_k^t;$$

they are well-defined since  $\phi_t^k$  and  $\chi_t^k$  vanish at points that belong to more than one set  $J_k^t$ . We have  $\mu(\phi) = \mu(\phi^g)$  by construction and the result follows.  $\square$

**Definition 4.15.** Let  $1 \leq t \leq n$  and  $x_t \in \mathbb{R}$ . A sequence  $\mathbf{x} = (x_0, \dots, x_t)$  is a *predecessor path* of  $x_t$  if there are indices  $(k_0, \dots, k_t)$  such that  $(x_{s-1}, x_s) \in (I_{k_s}^s, J_{k_s}^s)$  for some component (irreducible or diagonal) of  $\mathcal{M}(\mu_{s-1}, \mu_s)$ , for all  $1 \leq s \leq t$ . We write  $\mathbb{k}(\mathbf{x})$  for the (unique) associated sequence  $(k_0, \dots, k_t)$  followed by the path  $\mathbf{x}$  in the above sense, and  $\Psi_t^k(x_t)$  for the set of all predecessor paths with  $k_t = k$ .

These notions will be useful in the next step towards the closedness result, which is to “regularize” the concave moderators.

**Lemma 4.16.** *Let  $(\phi, H) \in \mathcal{D}_\mu^g(0)$ . There is a concave moderator  $\chi$  of  $\phi$  such that*

$$\phi_t^k + \chi_t^k - \chi_{t+1} \geq 0 \quad \text{on } J_k^t \quad \text{for all } t = 0, \dots, n, \quad k \geq 1. \quad (4.2)$$

As a consequence,

$$\sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k (\chi_t^k) \leq \mu(\phi).$$

*Proof.* Fix  $1 \leq t \leq n$  and let  $(I_k^t, J_k^t)$  be the domain of some component of  $\mathcal{M}(\mu_{t-1}, \mu_t)$ . We define  $\chi = (\chi_t^k)$  by  $\chi_0^0 = 0$  and

$$\chi_t^k(x_t) = \inf_{\mathbf{x} \in \Psi_t^k(x_t)} \left\{ \sum_{s=0}^{t-1} \phi_s^{\mathbb{k}_s(\mathbf{x})}(x_s) + (H \cdot \mathbf{x})_t \right\};$$



then  $\chi_t^k$  is concave on  $J_k^t$  as an infimum of affine functions. As  $\phi \in L^{c,g}(\mu)$ , we cannot have  $\sum_{s=0}^{t-1} \phi_s^{\mathbb{k}_s(\mathbf{x})}(x_s) = \infty$  for all  $\mathbf{x} \in \Psi_t^k(x_t)$ , so that  $\chi_t^k < \infty$  on  $J_k^t$ .

Next, we verify that  $\chi$  satisfies (4.2). For  $t = n$ , this follows directly from  $(\phi, H) \in \mathcal{D}_\mu^g(0)$ ; recall that  $\chi_{n+1} \equiv 0$ . For  $t < n$ , restricting the infimum in the definition of  $\chi$  to the set of paths  $\mathbf{x}$  with  $x_{t+1} = x_t \in I_{k'}^{t+1} \cap J_k^t$  yields

$$\begin{aligned} \chi_{t+1}(x_t) &= \chi_{t+1}^{k'}(x_t) = \inf_{\mathbf{x} \in \Psi_{t+1}^{k'}(x_t)} \left\{ \sum_{s=0}^t \phi_s^{\mathbb{k}_s(\mathbf{x})}(x_s) + (H \cdot \mathbf{x})_{t+1} \right\} \\ &\leq \inf_{\mathbf{x} \in \Psi_t^k(x_t)} \left\{ \sum_{s=0}^{t-1} \phi_s^{\mathbb{k}_s(\mathbf{x})}(x_s) + (H \cdot \mathbf{x})_t \right\} + \phi_t^k(x_t) \\ &= \chi_t^k(x_t) + \phi_t^k(x_t), \end{aligned}$$

and this implies (4.2) since  $\cup_{k' \geq 0} I_{k'}^{t+1} = \mathbb{R}$ .

The inequality (4.2) shows in particular that  $\chi_t^k \geq -\phi_t^k + \chi_{t+1}$  on  $I_k^t$ . Next, we show by induction that this implies  $\chi_t^k > -\infty$  on  $J_k^t$ . Indeed, as  $\chi_{n+1} \equiv 0$ , the latter is clear for  $t = n$ . From  $\phi \in L^{c,g}(\mu)$ , we see that the set  $\{\phi_t^k < \infty\}$  must be dense in  $J_k^t$  and contain any atoms of  $\mu_t^k$ . Using the induction hypothesis, i.e.  $\chi_{t+1} > -\infty$ , we have that  $\{\phi_t^k - \chi_{t+1} < \infty\}$  must be dense in  $J_k^t$  and contain any atoms of  $\mu_t^k$ . By the concavity of  $\chi_t^k$  we can then deduce that  $\chi_t^k > -\infty$  on  $J_k^t$  as desired.

Set  $\bar{\phi}_t^k := \phi_t^k + \chi_t^k - \chi_{t+1}|_{J_k^t}$  for  $0 \leq t \leq n$ ; then  $\bar{\phi}_t^k \geq 0$ . Moreover, choose an arbitrary  $P \in \mathcal{M}(\mu)$  with disintegration  $P = \mu_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_n$  for some stochastic kernels  $\kappa_t(x_0, \dots, x_{t-1}, dx_t)$ . From Lemma 4.8 we know that

$$\mu(\phi) = P \left[ \sum_{t=0}^n \phi_t^{\mathbb{k}_t(X)}(X_t) + (H \cdot X)_n \right] < \infty.$$

We can therefore apply Fubini's theorem for kernels as in the proof of Lemma 4.8 to the expression

$$\begin{aligned} 0 &\leq \sum_{t=0}^n \phi_t^{\mathbb{k}_t(\mathbf{x})}(x_t) + (H \cdot \mathbf{x})_n \\ &= \sum_{t=0}^n \bar{\phi}_t^{\mathbb{k}_t(\mathbf{x})}(x_t) + \sum_{t=1}^n \left( \chi_t(x_{t-1}) - \chi_t^{\mathbb{k}_t(\mathbf{x})}(x_t) \right) + (H \cdot \mathbf{x})_n \end{aligned}$$

and obtain

$$P \left[ \sum_{t=0}^n \phi_t^{\mathbb{K}_t(X)}(X_t) + (H \cdot X)_n \right] = \sum_{t=0}^n \sum_{k \geq 0} \mu_t^k(\bar{\phi}_t^k) + \sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t^k)$$

which shows that the right hand side is finite, and therefore  $\chi$  is a concave moderator for  $\phi$ . Finally, the second claim follows from  $\mu_t^k(\bar{\phi}_t^k) \geq 0$ .  $\square$

The last tool for our closedness result is a compactness property for concave functions in the one-step case; cf. [10, Proposition 5.5].

**Proposition 4.17.** *Let  $\mu \leq_c \nu$  be irreducible with domain  $(I, J)$  and let  $a \in I$  be the common barycenter of  $\mu$  and  $\nu$ . Let  $\chi_m : J \rightarrow \mathbb{R}$  be concave functions such that<sup>4</sup>*

$$\chi_m(a) = \chi'_m(a) = 0 \quad \text{and} \quad \sup_{m \geq 1} (\mu - \nu)(\chi_m) < \infty.$$

*There exists a subsequence  $\chi_{m_k}$  which converges pointwise on  $J$  to a concave function  $\chi : J \rightarrow \mathbb{R}$ , and  $(\mu - \nu)(\chi) \leq \liminf_k (\mu - \nu)(\chi_{m_k})$ .*

We are now ready to state and prove the analogue of Proposition 4.10 in the generalized dual.

**Proposition 4.18.** *Let  $f^m : \mathbb{R}^{n+1} \rightarrow [0, \infty)$ ,  $m \geq 1$  be a sequence of functions such that*

$$f^m \rightarrow f \quad \text{pointwise}$$

*and let  $(\phi^m, H^m) \in \mathcal{D}_\mu^g(f^m)$  be such that  $\sup_m \mu(\phi^m) < \infty$ . Then there exist  $(\phi, H) \in \mathcal{D}_\mu^g(f)$  with*

$$\mu(\phi) \leq \liminf_{m \rightarrow \infty} \mu(\phi^m).$$

*Proof.* Since  $(\phi^m, H^m) \in \mathcal{D}_\mu^g(f^m)$  and  $f^m \geq 0$ , we can introduce a sequence of concave moderators  $\chi_m$  as in Lemma 4.16. A normalization of  $(\phi^m, H^m)$  as in Remark 4.13 (i) and (ii) allows us to assume without loss of generality that  $\chi_{t,m}^0 \equiv 0$  and  $\chi_{t,m}^k(a_t^k) = (\chi_{t,m}^k)'(a_t^k) = 0$ , where  $a_t^k$  is the barycenter of  $\mu_t^k$ —this modification is the main merit of lifting to the generalized dual space. By passing to a subsequence as in Proposition 4.17 for each component and using a diagonal argument, we obtain pointwise limits  $\chi_t^k : J_k^t \rightarrow \mathbb{R}$  for  $\chi_{t,m}^k$  after passing to another subsequence.

---

<sup>4</sup>To be specific, let us convene that  $\chi'_m$  is the left derivative—this is not important here.

Since  $\phi_{t,m}^k + \chi_{t,m}^k - \chi_{t+1,m} \geq 0$  on  $J_k^t$  and  $\chi_{t,m}^k \rightarrow \chi_t^k$  as well as  $\chi_{t+1,m} \rightarrow \chi_{t+1}$ , we can apply Komlos' lemma to find  $\tilde{\phi}_{t,m}^k \in \text{conv}\{\phi_{t,m}^k, \phi_{t,m+1}^k, \dots\}$  which converge  $\mu_t^k$ -a.s. for  $0 \leq t \leq n$ . We may assume without loss of generality that  $\tilde{\phi}_{t,m}^k = \phi_{t,m}^k$ . Thus, we can set

$$\phi_t^k := \limsup \phi_{t,m}^k \quad \text{on } J_k^t$$

to obtain

$$\phi_{t,m}^k \rightarrow \phi_t^k \quad \mu_t^k\text{-a.s.} \quad \text{and} \quad \phi_t^k + \chi_t^k - \chi_{t+1} \geq 0 \quad \text{on } J_k^t.$$

We can now apply Fatou's lemma and Proposition 4.17 to deduce that

$$\begin{aligned} \mu(\phi) &= \sum_{t=0}^n \sum_{k \geq 0} \mu_t^k(\phi_t^k + \chi_t^k - \chi_{t+1}) + \sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t^k) \\ &\leq \sum_{t=0}^n \sum_{k \geq 0} \liminf \mu_t^k(\phi_{t,m}^k + \chi_{t,m}^k - \chi_{t+1,m}) \\ &\quad + \sum_{t=1}^n \sum_{k \geq 1} \liminf (\mu_{t-1} - \mu_t)^k(\chi_{t,m}^k) \\ &\leq \liminf \left[ \sum_{t=0}^n \sum_{k \geq 0} \mu_t^k(\phi_{t,m}^k + \chi_{t,m}^k - \chi_{t+1,m}) + \sum_{t=1}^n \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k(\chi_{t,m}^k) \right] \\ &= \liminf \mu(\phi^m) < \infty. \end{aligned}$$

In particular, we see that  $\phi \in L^{c,g}(\mu)$  with concave moderator  $\chi$ .

It remains to construct the predictable process  $H = (H_1, \dots, H_n)$ . With a mild abuse of notation, we shall identify  $H_t(x_0, \dots, x_n)$  with the corresponding function of  $(x_0, \dots, x_{t-1})$  in this proof.

We first define for each  $\mathbf{k} = (k_0, \dots, k_t)$  and  $\mathbf{x} = (x_0, \dots, x_t)$  such that  $\mathbf{k} = \mathbb{k}(\mathbf{x})$ , the functions  $G_{t,m}^{\mathbf{k}}$  and  $G_t^{\mathbf{k}}$  by

$$\begin{aligned} G_{t,m}^{\mathbf{k}}(\mathbf{x}) &:= \sum_{s=0}^t \phi_{s,m}^{k_s}(x_s) + \sum_{s=1}^t H_{s,m}(x_0, \dots, x_{s-1}) \cdot (x_s - x_{s-1}), \\ G_t^{\mathbf{k}}(\mathbf{x}) &:= \liminf G_{t,m}^{\mathbf{k}}(\mathbf{x}). \end{aligned}$$

Given  $\mathbf{k} = (k_0, \dots, k_t)$ , we write  $\mathbf{k}' = (k_0, \dots, k_{t-1})$ . We claim that there exists an  $\mathbb{F}$ -predictable process  $H$  such that for all  $1 \leq t \leq n$ ,

$$G_{t-1}^{\mathbf{k}'}(x_0, \dots, x_{t-1}) + \phi_t^{k_t}(x_t) + H_t(x_0, \dots, x_{t-1}) \cdot (x_t - x_{t-1}) \geq G_t^{\mathbf{k}}(x_0, \dots, x_t). \quad (4.3)$$

Once this is established, the proposition follows by induction since  $G_0^{(0)}(x_0) = \phi_0(x_0)$  and  $G_n^{\mathbf{k}}(x_0, \dots, x_n) \geq f(x_0, \dots, x_n)$ .

To prove the claim, write  $g^{\text{conc}}$  for the concave hull of a function  $g$  and observe that

$$\begin{aligned} \liminf [G_{t-1,m}^{\mathbf{k}'}(x_0, \dots, x_{t-1}) + H_{t,m}(x_0, \dots, x_{t-1}) \cdot (x_t - x_{t-1})] \\ \geq \liminf [(G_{t,m}^{\mathbf{k}}(x_0, \dots, x_{t-1}, \cdot) - \phi_{t,m}^{k_t}(\cdot))^{\text{conc}}(x_t)] \\ \geq [\liminf (G_{t,m}^{\mathbf{k}}(x_0, \dots, x_{t-1}, \cdot) - \phi_{t,m}^{k_t}(\cdot))^{\text{conc}}(x_t)] \\ = [G_t^{\mathbf{k}}(x_0, \dots, x_{t-1}, \cdot) - \phi_t^{k_t}(\cdot)]^{\text{conc}}(x_t) \\ =: \hat{\phi}_t^{\mathbf{k}}(x_0, \dots, x_{t-1}, x_t). \end{aligned}$$

By construction,  $\hat{\phi}_t^{\mathbf{k}}$  is concave in the last variable and satisfies

$$G_{t-1}^{\mathbf{k}'}(x_0, \dots, x_{t-1}) \geq \hat{\phi}_t^{\mathbf{k}}(x_0, \dots, x_{t-1}, x_{t-1}).$$

Let  $\partial_t \hat{\phi}_t^{\mathbf{k}}$  denote the right partial derivative in the last variable and set

$$H_t^{\mathbf{k}}(x_0, \dots, x_{t-1}) := \partial_t \hat{\phi}_t^{\mathbf{k}}(x_0, \dots, x_{t-1}, x_{t-1});$$

then we have

$$\begin{aligned} G_{t-1}^{\mathbf{k}'}(x_0, \dots, x_{t-1}) + H_t^{\mathbf{k}}(x_0, \dots, x_{t-1}) \cdot (x_t - x_{t-1}) \\ \geq \hat{\phi}_t^{\mathbf{k}}(x_0, \dots, x_{t-1}, x_{t-1}) + H_t^{\mathbf{k}}(x_0, \dots, x_{t-1}) \cdot (x_t - x_{t-1}) \\ \geq \hat{\phi}_t^{\mathbf{k}}(x_0, \dots, x_{t-1}, x_t) \\ \geq G_t^{\mathbf{k}}(x_0, \dots, x_t) - \phi_t^{k_t}(x_t). \end{aligned}$$

Finally, for any  $(x_0, \dots, x_{t-1}) \in \mathbb{R}^t$ , we define  $H_t(x_0, \dots, x_{t-1})$  as

$$\begin{cases} H_t^{\mathbf{k}}(x_0, \dots, x_{t-1}), & \text{if } \mathbf{k} = \mathbb{k}(x_0, \dots, x_{t-1}, x_t) \text{ for some } x_t \in \mathbb{R} \\ 0, & \text{otherwise;} \end{cases}$$

this is well-defined since  $\mathbb{k}(x_0, \dots, x_t)$  depends only on  $(x_0, \dots, x_{t-1})$ . The predictable process  $H$  satisfies (4.3) and thus the proof is complete.  $\square$

*Proof of Proposition 4.10.* In view of Remark 4.13 (iii) and Proposition 4.14, the result follows from Proposition 4.18.  $\square$

## 5 Duality Theorem and Monotonicity Principle

The first goal of this section is a duality result for the multistep martingale transport problem; it establishes the absence of a duality gap and the existence of optimizers in the dual problem. (As is well known, an optimizer for the primal problem only exists under additional conditions, such as continuity of  $f$ .) The second goal is a monotonicity principle describing the geometry of optimal transports; it will be a consequence of the duality result.

As above, we consider a fixed vector  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  of marginals in convex order. The primal and dual problems as defined follows.

**Definition 5.1.** Let  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$ . The *primal problem* is

$$\mathbf{S}_{\boldsymbol{\mu}}(f) := \sup_{P \in \mathcal{M}(\boldsymbol{\mu})} P(f) \in [0, \infty],$$

where  $P(f)$  refers to the outer integral if  $f$  is not measurable. The *dual problem* is

$$\mathbf{I}_{\boldsymbol{\mu}}(f) := \inf_{(\phi, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)} \boldsymbol{\mu}(\phi) \in [0, \infty].$$

We recall that a function  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$  is called *upper semianalytic* if the sets  $\{f \geq c\}$  are analytic for all  $c \in \mathbb{R}$ , where a subset of  $\mathbb{R}^{n+1}$  is called analytic if it is the image of a Borel subset of a Polish space under a Borel mapping. Any Borel function is upper semianalytic and any upper semianalytic function is universally measurable; we refer to [11, Chapter 7] for background. The following is the announced duality result.

**Theorem 5.2 (Duality).** Let  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$ .

- (i) If  $f$  is upper semianalytic, then  $\mathbf{S}_{\boldsymbol{\mu}}(f) = \mathbf{I}_{\boldsymbol{\mu}}(f) \in [0, \infty]$ .
- (ii) If  $\mathbf{I}_{\boldsymbol{\mu}}(f) < \infty$ , there exists a dual optimizer  $(\phi, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$ .

*Proof.* Given our preceding results, much of the proof follows the lines of the corresponding result for the one-step case in [10, Theorem 6.2]; therefore, we shall be brief. We mention that the present theorem is slightly more general than the cited one in terms of the measurability condition on  $f$ ; this is due to the global proof given here.

*Step 1.* Using Lemma 4.8 we see that  $\mathbf{S}_{\boldsymbol{\mu}}(f) \leq \mathbf{I}_{\boldsymbol{\mu}}(f)$  holds for all upper semicontinuous  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$ .

*Step 2.* Using the de la Vallée–Poussin theorem and our assumption that the marginals have a finite first moment, there exist increasing, superlinearly

growing functions  $\zeta_{\mu_t} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that  $x \mapsto \zeta_{\mu_t}(|x|)$  is  $\mu_t$ -integrable for all  $0 \leq t \leq n$ . Define

$$\zeta(x_0, \dots, x_n) := 1 + \sum_{t=0}^n \zeta_{\mu_t}(|x_t|)$$

and let  $C_\zeta$  be the vector space of all continuous functions  $f$  such that  $f/\zeta$  vanishes at infinity. Then, a Hahn–Banach separation argument can be used to show that  $\mathbf{S}_\mu(f) \geq \mathbf{I}_\mu(f)$  holds for all  $f \in C_\zeta$ .

*Step 3.* Let  $f$  be bounded and upper semicontinuous; then there exists a sequence of bounded continuous functions  $f^m \in C_b(\mathbb{R}^{n+1})$  which decrease to  $f$  pointwise. As  $C_b(\mathbb{R}^{n+1}) \subseteq C_\zeta$ , we have  $\mathbf{S}_\mu(f^m) = \mathbf{I}_\mu(f^m)$  for all  $m$  by the first two steps.

Let  $\mathcal{U}$  be the set of all bounded, nonnegative, upper semicontinuous functions on  $\mathbb{R}^{n+1}$ . We recall that a map  $\mathbf{C} : [0, \infty]^{\mathbb{R}^{n+1}} \rightarrow [0, \infty]$  is called a  $\mathcal{U}$ -capacity if it is monotone, sequentially continuous upwards on  $[0, \infty]^{\mathbb{R}^{n+1}}$  and sequentially continuous downwards on  $\mathcal{U}$ . The functional  $f \mapsto \mathbf{S}_\mu(f)$  is a  $\mathcal{U}$ -capacity; this follows from the weak compactness of  $\mathcal{M}(\mu)$  and the arguments in [32, Propositions 1.21, 1.26].

It follows that  $\mathbf{S}_\mu(f^m) \rightarrow \mathbf{S}_\mu(f)$ . By the monotonicity of  $f \mapsto \mathbf{I}_\mu(f)$  and Step 1 we obtain

$$\mathbf{I}_\mu(f) \leq \lim \mathbf{I}_\mu(f^m) = \lim \mathbf{S}_\mu(f^m) = \mathbf{S}_\mu(f) \leq \mathbf{I}_\mu(f).$$

*Step 4.* Since  $\mathbf{S}_\mu = \mathbf{I}_\mu$  on  $\mathcal{U}$  by Step 3,  $\mathbf{I}_\mu$  is sequentially downward continuous on  $\mathcal{U}$  like  $\mathbf{S}_\mu$ . On the other hand, Proposition 4.10 implies that it is sequentially upwards continuous on  $[0, \infty]^{\mathbb{R}^{n+1}}$ . As a result,  $\mathbf{I}_\mu$  is a  $\mathcal{U}$ -capacity.

*Step 5.* Let  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$  be upper semianalytic. For any  $\mathcal{U}$ -capacity  $\mathbf{C}$ , Choquet's capacitability theorem shows that

$$\mathbf{C}(f) = \sup\{\mathbf{C}(g) : g \in \mathcal{U}, g \leq f\}.$$

As  $\mathbf{S}_\mu$  and  $\mathbf{I}_\mu$  are  $\mathcal{U}$ -capacities that coincide on  $\mathcal{U}$ , it follows that  $\mathbf{S}_\mu(f) = \mathbf{I}_\mu(f)$ . This completes the proof of (i).

*Step 6.* To see that the infimum  $\mathbf{I}_\mu(f)$  is attained if it is finite, we merely need to apply Proposition 4.10 with the constant sequence  $f^m = f$ .  $\square$

As announced above, we can easily relax the lower bound on  $f$ .

**Remark 5.3.** Let  $f : \mathbb{R}^{n+1} \rightarrow (-\infty, \infty]$  and suppose there exist  $\phi \in \prod_{t=0}^n L^1(\mu_t)$  and a predictable process  $H$  such that

$$f \geq \sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \quad \text{on } \mathcal{V}.$$

Then we can apply Theorem 5.2 to  $[f - \sum_{t=0}^n \phi_t(X_t) - (H \cdot X)_n]^+$  and obtain the analogue of its assertion for  $f$ .

The duality result gives rise to a monotonicity principle describing the support of optimal martingale transports, in the spirit of the cyclical monotonicity condition from classical transport theory. The following generalizes the results of [8, Lemma 1.11] and [10, Corollary 7.8] for the one-step martingale transport problem.

**Theorem 5.4** (Monotonicity Principle). *Let  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$  be Borel and suppose that  $\mathbf{S}_\mu(f) < \infty$ . There exists a Borel set  $\Gamma \subseteq \mathbb{R}^{n+1}$  with the following properties.*

- (i) *A measure  $P \in \mathcal{M}(\mu)$  is concentrated on  $\Gamma$  if and only if it is optimal for  $\mathbf{S}_\mu(f)$ .*
- (ii) *Let  $\bar{\mu} = (\bar{\mu}_0, \dots, \bar{\mu}_n)$  be another vector of marginals in convex order. If  $\bar{P} \in \mathcal{M}(\bar{\mu})$  is concentrated on  $\Gamma$ , then  $\bar{P}$  is optimal for  $\mathbf{S}_{\bar{\mu}}(f)$ .*

Indeed, if  $(\phi, H) \in \mathcal{D}_\mu(f)$  is an optimizer for  $\mathbf{I}_\mu(f)$ , then we can take

$$\Gamma := \left\{ \sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n = f \right\} \cap \mathcal{V}.$$

*Proof.* As  $\mathbf{S}_\mu(f) < \infty$ , Theorem 5.2 shows that  $\mathbf{I}_\mu(f) = \mathbf{S}_\mu(f) < \infty$  and that there exists a dual optimizer  $(\phi, H) \in \mathcal{D}_\mu(f)$ . In particular, we can define  $\Gamma$  as above.

(i) As  $0 \leq f$  and  $P(f) \leq \mathbf{S}_\mu(f) < \infty$  for all  $P \in \mathcal{M}(\mu)$ , we see that  $f$  is  $P$ -integrable for all  $P \in \mathcal{M}(\mu)$ . Since  $\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \geq 0$  on the effective domain  $\mathcal{V}$ , and  $P[\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n] = \mu(\phi) = \mathbf{I}_\mu(f) < \infty$  by Lemma 4.8, we also obtain the  $P$ -integrability of  $\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n$ . In particular,

$$0 \leq P \left[ \sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n - f \right] = \mu(\phi) - P(f) = \mathbf{S}_\mu(f) - P(f)$$

and equality holds if and only if  $P$  is concentrated on  $\Gamma$ .

(ii) We may assume that  $\bar{P}$  is a probability measure with  $\bar{P}(f) < \infty$ . As a first step, we show that the effective domain  $\bar{\mathcal{V}}$  of  $\mathcal{M}(\bar{\mu})$  is a subset of the effective domain  $\mathcal{V}$  of  $\mathcal{M}(\mu)$ . To that end, it is sufficient to show that if  $1 \leq t \leq n$  and  $x \in \mathbb{R}$  are such that  $u_{\mu_{t-1}}(x) = u_{\mu_t}(x)$ , then  $u_{\bar{\mu}_{t-1}}(x) = u_{\bar{\mu}_t}(x)$ ; cf. Proposition 2.3. Indeed, for such  $t$  and  $x$ , our assumption that  $\Gamma \subseteq \mathcal{V}$  implies

$$\Gamma \subseteq (X_t, X_{t-1})^{-1}((-\infty, x]^2 \cup [x, \infty)^2).$$

Using also that  $\mathbb{E}^{\bar{P}}[X_t | \mathfrak{F}_{t-1}] = X_{t-1}$  and that  $\bar{P}$  is concentrated on  $\Gamma$ ,

$$\begin{aligned} u_{\bar{\mu}_{t-1}}(x) &= \mathbb{E}^{\bar{P}}[|X_{t-1} - x|] \\ &= \mathbb{E}^{\bar{P}}[(X_{t-1} - x)\mathbf{1}_{X_{t-1} \geq x}] + \mathbb{E}^{\bar{P}}[(x - X_{t-1})\mathbf{1}_{X_{t-1} \leq x}] \\ &= \mathbb{E}^{\bar{P}}[(X_t - x)\mathbf{1}_{X_{t-1} \geq x}] + \mathbb{E}^{\bar{P}}[(x - X_t)\mathbf{1}_{X_{t-1} \leq x}] \\ &= \mathbb{E}^{\bar{P}}[|X_t - x|] = u_{\bar{\mu}_t}(x) \end{aligned}$$

as desired, and we have shown that  $\bar{\mathcal{V}} \subseteq \mathcal{V}$ .

In view of that inclusion, the inequality  $\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \geq f$  holds on  $\bar{\mathcal{V}}$ . Since  $\bar{P}$  is concentrated on  $\Gamma$ ,

$$\bar{P} \left[ \sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \right] = \bar{P}(f) < \infty.$$

Using this fact as a substitute for the assumption that  $(\phi, H) \in \mathcal{D}_{\mu}^g(0)$ , we may follow the arguments in the proof of Lemma 4.16 to construct a moderator  $\chi$  and establish that  $(\phi, H) \in \mathcal{D}_{\mu}^g(f)$ , where we are implicitly using the embedding detailed in Remark 4.13 (iii). Then, we can modify  $\chi$  as in the proof of Proposition 4.14 to see that  $(\phi, H) \in \mathcal{D}_{\mu}(f)$ . As a result, we may apply Lemma 4.8 to obtain that

$$\bar{P}(f) = \bar{P} \left[ \sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \right] = \bar{\mu}(\phi),$$

whereas for any other  $P' \in \mathcal{M}(\bar{\mu})$  we have

$$P'(f) \leq P' \left[ \sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \right] = \bar{\mu}(\phi) = \bar{P}(f).$$

This shows that  $\bar{P} \in \mathcal{M}(\bar{\mu})$  is optimal.  $\square$



## 6 Left-Monotone Transports

In this section we define left-monotone transports through a shadow property and prove their existence.

### 6.1 Preliminaries

Before moving on to the  $n$ -step case, we recall the essential definitions and results regarding the one-step version of the left-monotone transport (also called the Left-Curtain coupling). The first notion is the so-called shadow, and it will be useful to define it for measures  $\mu \leq_{pc} \nu$  in *positive convex order*, meaning that  $\mu(\phi) \leq \nu(\phi)$  for any nonnegative convex function  $\phi$ . Clearly, this order is weaker than the convex order  $\mu \leq_c \nu$ , and it is worth noting that  $\mu$  may have a smaller mass than  $\nu$ . The following is the result of [8, Lemma 4.6].

**Lemma 6.1.** *Let  $\mu \leq_{pc} \nu$ . Then the set*

$$\llbracket \mu; \nu \rrbracket := \{ \theta : \mu \leq_c \theta \leq \nu \}$$

*is non-empty and contains a unique least element  $\mathcal{S}^\nu(\mu)$  for the convex order:*

$$\mathcal{S}^\nu(\mu) \leq_c \theta \text{ for all } \theta \in \llbracket \mu; \nu \rrbracket.$$

*The measure  $\mathcal{S}^\nu(\mu)$  is called the shadow of  $\mu$  in  $\nu$ .*

It will be useful to have the following picture in mind: if  $\mu$  is a Dirac measure, its shadow in  $\nu$  is a measure  $\theta$  of equal mass and barycenter, chosen such as to have minimal variance subject to the constraint  $\theta \leq \nu$ .

The second notion is a class of reward functions.

**Definition 6.2.** A Borel function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is called *second-order Spence–Mirrlees* if  $y \mapsto f(x', y) - f(x, y)$  is strictly convex for any  $x < x'$ .

We note that if  $f$  is sufficiently differentiable, this can be expressed as the cross-derivative condition  $f_{xyy} > 0$  which has also been called the martingale Spence–Mirrlees condition, in analogy to the classical Spence–Mirrlees condition  $f_{xy} > 0$ .

In the one-step case, the left-monotone transport is unique and can be characterized as follows; cf. [8, Theorems 4.18, 4.21, 6.1] where this transport is called the Left-Curtain coupling, as well as [35, Theorem 1.2] for the third equivalence in the stated generality.

**Proposition 6.3.** *Let  $\mu \leq_c \nu$  and  $P \in \mathcal{M}(\mu, \nu)$ . The following are equivalent:*

(i) *For all  $x \in \mathbb{R}$  and  $A \in \mathfrak{B}(\mathbb{R})$ ,*

$$P[(-\infty, x] \times A] = \mathcal{S}^\nu(\mu|_{(-\infty, x]})(A).$$

(ii)  *$P$  is concentrated on a Borel set  $\Gamma \subseteq \mathbb{R}^2$  satisfying*

$$(x, y^-), (x, y^+), (x', y') \in \Gamma, \quad x < x' \quad \Rightarrow \quad y' \notin (y^-, y^+).$$

(iii) *Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be second-order Spence–Mirrlees and assume that there exist functions  $a \in L^1(\mu)$ ,  $b \in L^1(\nu)$  such that  $|f(x, y)| \leq a(x) + b(y)$ . Then  $P$  is an optimizer for  $\mathbf{S}_{\mu, \nu}(f)$ .*

*There exists a unique measure  $\bar{P} \in \mathcal{M}(\mu, \nu)$  satisfying (i)–(iii), and  $\bar{P}$  is called the (one-step) left-monotone transport.*

If  $\mu$  is a discrete measure, the characterization in (i) can be understood as follows: the left-monotone transport  $\bar{P}$  processes the atoms of  $\mu$  from left to right, mapping each one of them to its shadow in the remaining target measure.

Next, we record two more results about shadows that will be used below. The first one, cited from [9, Theorem 3.1], generalizes the above idea in the sense that the atoms are still mapped to their shadows but can be processed in any given order; in the general (non-discrete) case, such an order is defined by a coupling  $\pi$  from the uniform measure to  $\mu$ .

**Proposition 6.4.** *Let  $\mu \leq_c \nu$  and  $\pi \in \Pi(\lambda, \mu)$  where  $\lambda$  denotes the Lebesgue measure on  $[0, 1]$ . Then there exists a unique measure  $Q \in \Pi(\lambda, \mu, \nu)$  on  $\mathbb{R}^3$  such that  $Q \circ (X_0, X_1)^{-1} = \pi$  and*

$$Q|_{[0, s] \times \mathbb{R} \times \mathbb{R}} \circ (X_1, X_2)^{-1} \in \mathcal{M}(\pi_s, \mathcal{S}^\nu(\pi_s)), \quad s \in \mathbb{R},$$

where  $\pi_s := \pi|_{[0, s] \times \mathbb{R}} \circ (X_1)^{-1}$ .

We shall also need the following facts about shadows.

**Lemma 6.5.** (i) *Let  $\mu_1, \mu_2, \nu$  be finite measures satisfying  $\mu_1 + \mu_2 \leq_{pc} \nu$ . Then  $\mu_2 \leq_{pc} \nu - \mathcal{S}^\nu(\mu_1)$  and  $\mathcal{S}^\nu(\mu_1 + \mu_2) = \mathcal{S}^\nu(\mu_1) + \mathcal{S}^{\nu - \mathcal{S}^\nu(\mu_1)}(\mu_2)$ .*

(ii) *Let  $\mu, \nu_1, \nu_2$  be finite measures such that  $\mu \leq_{pc} \nu_1 \leq_c \nu_2$ . Then, it follows that  $\mathcal{S}^{\nu_1}(\mu) \leq_{pc} \nu_2$ . Moreover,  $\mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_1}(\mu)) = \mathcal{S}^{\nu_2}(\mu)$  if and only if  $\mathcal{S}^{\nu_1}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mu)$ .*

*Proof.* Part (i) is [8, Theorem 4.8]. To obtain the first statement in (ii), we observe that  $\mathcal{S}^{\nu_1}(\mu) \leq \nu_1 \leq_c \nu_2$  and hence

$$\mathcal{S}^{\nu_1}(\mu)(\phi) \leq \nu_1(\phi) \leq \nu_2(\phi)$$

for any nonnegative convex function  $\phi$ . Turning to the second statement, the “only if” implication follows directly from the definition of the shadow in Lemma 6.1. To show the reverse implication, suppose that  $\mathcal{S}^{\nu_1}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mu)$ . Then, we have

$$\mu \leq_c \mathcal{S}^{\nu_1}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_1}(\mu)) \leq \nu_2 \quad \text{and} \quad \mathcal{S}^{\nu_1}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mu) \leq \nu_2.$$

These inequalities imply that

$$\mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_1}(\mu)) \in \llbracket \mu; \nu_2 \rrbracket \quad \text{and} \quad \mathcal{S}^{\nu_2}(\mu) \in \llbracket \mathcal{S}^{\nu_1}(\mu); \nu_2 \rrbracket,$$

and now the minimality property of the shadow shows that

$$\mathcal{S}^{\nu_2}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_1}(\mu)) \leq_c \mathcal{S}^{\nu_2}(\mu)$$

as desired.  $\square$

## 6.2 Construction of a Multistep Left-Monotone Transport

Our next goal is to define and construct a multistep left-monotone transport. The following concept will be crucial.

**Definition 6.6.** Let  $\mu_0 \leq_{pc} \mu_1 \leq_c \cdots \leq_c \mu_n$ . For  $1 \leq t \leq n$ , the *obstructed shadow of  $\mu_0$  in  $\mu_t$  through  $\mu_1, \dots, \mu_{t-1}$*  is iteratively defined by

$$\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0) := \mathcal{S}^{\mu_t}(\mathcal{S}^{\mu_1, \dots, \mu_{t-1}}(\mu_0)).$$

The obstructed shadow is well-defined due to Lemma 6.5 (ii). An alternative definition is provided by the following characterization.

**Lemma 6.7.** Let  $\mu_0 \leq_{pc} \mu_1 \leq_c \cdots \leq_c \mu_n$  and  $1 \leq t \leq n$ . Then  $\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0)$  is the unique least element of the set

$$\llbracket \mu_0; \mu_t \rrbracket^{\mu_1, \dots, \mu_{t-1}} := \{ \theta_t \leq \mu_t : \exists \theta_s \leq \mu_s, 1 \leq s \leq t-1, \mu_0 \leq_c \theta_1 \leq_c \cdots \leq_c \theta_t \}$$

for the convex order; that is,  $\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0) \leq_c \theta$  for all elements  $\theta$ .

*Proof.* For  $t = 1$  this holds by the definition of the shadow in Lemma 6.1. For  $t > 1$ , we inductively assume that  $\mathcal{S}^{\mu_1, \dots, \mu_{t-1}}(\mu_0)$  is the least element of  $\llbracket \mu_0; \mu_{t-1} \rrbracket^{\mu_1, \dots, \mu_{t-2}}$ . Consider an arbitrary element  $\theta_t \in \llbracket \mu_0; \mu_t \rrbracket^{\mu_1, \dots, \mu_{t-1}}$  and fix some

$$\mu_0 \leq_c \theta_1 \leq_c \dots \leq_c \theta_{t-1} \leq_c \theta_t \quad \text{with} \quad \theta_s \leq \mu_s, \quad 1 \leq s \leq t-1.$$

Then,  $\theta_{t-1} \in \llbracket \mu_0; \mu_{t-1} \rrbracket^{\mu_1, \dots, \mu_{t-2}}$  and in particular  $\mathcal{S}^{\mu_1, \dots, \mu_{t-1}}(\mu_0) \leq_c \theta_{t-1}$ . Recall that  $\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0)$  is defined as the least element of

$$\begin{aligned} \llbracket \mathcal{S}^{\mu_1, \dots, \mu_{t-1}}(\mu_0); \mu_t \rrbracket &= \{\theta \leq \mu_t : \mathcal{S}^{\mu_1, \dots, \mu_{t-1}}(\mu_0) \leq_c \theta\} \\ &\supseteq \{\theta \leq \mu_t : \theta_{t-1} \leq_c \theta\} \ni \theta_t. \end{aligned}$$

Hence,  $\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0) \leq_c \theta_t$ , and as  $\theta_t \in \llbracket \mu_0; \mu_t \rrbracket^{\mu_1, \dots, \mu_{t-1}}$  was arbitrary, this shows that  $\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0)$  is a least element of  $\llbracket \mu_0; \mu_t \rrbracket^{\mu_1, \dots, \mu_{t-1}}$ . The uniqueness of the least element follows from the general fact that  $\theta_t^1 \leq_c \theta_t^2$  and  $\theta_t^2 \leq_c \theta_t^1$  imply  $\theta_t^1 = \theta_t^2$ .  $\square$

We can now state the main result of this section.

**Theorem 6.8.** *Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order. Then there exists  $P \in \mathcal{M}(\boldsymbol{\mu})$  such that the bivariate projections  $P_{0t} := P \circ (X_0, X_t)^{-1}$  satisfy*

$$P_{0t}[(-\infty, x] \times A] = \mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0|_{(-\infty, x]})(A) \quad \text{for } x \in \mathbb{R}, A \in \mathfrak{B}(\mathbb{R}),$$

for all  $1 \leq t \leq n$ . Any such  $P \in \mathcal{M}(\boldsymbol{\mu})$  is called a left-monotone transport.

We observe that an  $n$ -step left-monotone transport is defined purely in terms of its bivariate projections  $P \circ (X_0, X_t)^{-1}$ . In the one-step case, this completely determines the transport. For  $n > 1$ , we shall see that there can be multiple (and then infinitely many) left-monotone transports; in fact, they form a convex compact set. This will be discussed in more detail in Section 8, where it will also be shown that uniqueness does hold if  $\mu_0$  is atomless.

*Proof of Theorem 6.8. Step 1.* We first construct measures  $\pi_t \in \Pi(\lambda, \mu_t)$ ,  $0 \leq t \leq n$  such that

$$\pi_t|_{[0, \mu_0((-\infty, x])] \times \mathbb{R}} \circ X_1^{-1} = \mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0|_{(-\infty, x]})$$

for all  $x \in \mathbb{R}$ , as well as measures  $Q_t \in \Pi(\lambda, \mu_{t-1}, \mu_t)$ ,  $1 \leq t \leq n$  such that

$$\begin{aligned} Q_t|_{[0, \mu_0((-\infty, x])] \times \mathbb{R} \times \mathbb{R}} \circ (X_1, X_2)^{-1} &\in \\ \mathcal{M}(\mathcal{S}^{\mu_1, \dots, \mu_{t-1}}(\mu_0|_{(-\infty, x]}), \mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0|_{(-\infty, x]})) & \end{aligned}$$

for all  $x \in \mathbb{R}$ . Indeed, for  $t = 0$ , we take  $\pi_0 \in \Pi(\lambda, \mu_0)$  to be the quantile<sup>5</sup> coupling. Then, applying Proposition 6.4 to  $\pi_0$  yields the measure  $Q_1$ , and we can define  $\pi_1 := Q_1 \circ (X_0, X_2)^{-1}$ . Proceeding inductively, applying Proposition 6.4 to  $\pi_{t-1}$  yields  $Q_t$  which in turn allows us to define  $\pi_t := Q_t \circ (X_0, X_2)^{-1}$ .

*Step 2.* For  $1 \leq t \leq n$ , consider a disintegration  $Q_t = \pi_{t-1} \otimes \kappa_t$  of  $Q_t$ . By the martingale property of  $Q_t$ , we may choose  $\kappa_t(s, x_{t-1}, dx_t)$  to be a martingale kernel; that is,

$$\int x_t \kappa_t(s, x_{t-1}, dx_t) = x_{t-1}$$

holds for all  $(s, x_{t-1}) \in \mathbb{R}^2$ . We now define a measure  $\pi \in \Pi(\lambda, \mu_0, \dots, \mu_n)$  on  $\mathbb{R}^{n+2}$  via

$$\pi = \pi_0 \otimes \kappa_1 \otimes \dots \otimes \kappa_n.$$

Then,  $\pi$  satisfies

$$\pi \circ (X_0, X_t)^{-1} = \pi_{t-1} \quad \text{and} \quad \pi \circ (X_0, X_t, X_{t+1})^{-1} = Q_t$$

for  $1 \leq t \leq n$ , and setting  $P = \pi \circ (X_1, \dots, X_{n+1})^{-1}$  yields the theorem.  $\square$

The following result studies the bivariate projections  $P_{0t}$  of a left-monotone transport and shows in particular that  $P_{0t}$  may differ from the Left-Curtain coupling [8] in  $\mathcal{M}(\mu_0, \mu_t)$ .

**Proposition 6.9.** *Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order and let  $P \in \mathcal{M}(\boldsymbol{\mu})$  be a left-monotone transport. The following are equivalent:*

- (i) *The bivariate projection  $P_{0t} = P \circ (X_0, X_t)^{-1} \in \mathcal{M}(\mu_0, \mu_t)$  is left-monotone for all  $1 \leq t \leq n$ .*
- (ii) *The marginals  $\boldsymbol{\mu}$  satisfy*

$$\mathcal{S}^{\mu_1}(\mu_0|_{(-\infty, x]}) \leq_c \dots \leq_c \mathcal{S}^{\mu_n}(\mu_0|_{(-\infty, x]}) \quad \text{for all } x \in \mathbb{R}. \quad (6.1)$$

*Proof.* Given  $\mu \leq \mu_0$ , an iterative application of Lemma 6.5 (ii) shows that the obstructed shadows coincide with the ordinary shadows, i.e.  $\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu) = \mathcal{S}^{\mu_t}(\mu)$  for  $1 \leq t \leq n$ , if and only if  $\mathcal{S}^{\mu_1}(\mu) \leq_c \dots \leq_c \mathcal{S}^{\mu_n}(\mu)$ . The proposition follows by applying this observation to  $\mu = \mu_0|_{(-\infty, x]}$ .  $\square$

The following example illustrates the proposition and shows that (6.1) may indeed fail.

---

<sup>5</sup>The quantile coupling (or Fréchet–Hoeffding coupling) is given by the law of  $(F_\lambda^{-1}, F_{\mu_0}^{-1})$  under  $\lambda$ , where  $F_{\mu_0}^{-1}$  is the inverse c.d.f. of  $\mu_0$ .

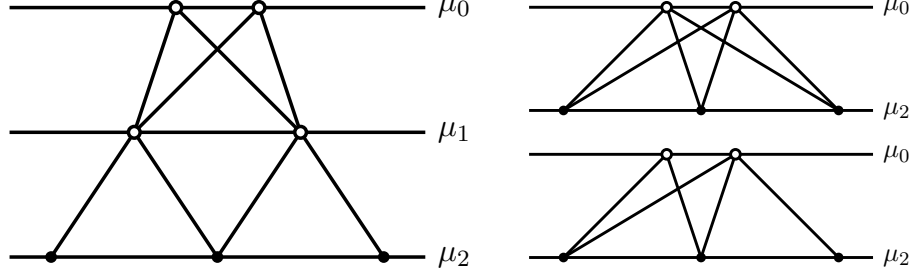


Figure 3: The left panel shows the support of the left-monotone transport  $P$  from Example 6.10. The right panel shows the support of  $P_{02}$  (top) and the support of the left-monotone transport in  $\mathcal{M}(\mu_0, \mu_2)$  (bottom).

**Example 6.10.** Consider the marginals

$$\mu_0 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1, \quad \mu_1 = \frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_2, \quad \mu_2 = \frac{1}{4}\delta_{-4} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_4.$$

Then the set  $\mathcal{M}(\mu)$  consists of a single transport  $P$ ; cf. the left panel of Figure 3. Thus,  $P$  is necessarily left-monotone. Similarly,  $P_{01} = P \circ (X_0, X_1)^{-1}$  is the unique element of  $\mathcal{M}(\mu_0, \mu_1)$ . However,  $P_{02} = P \circ (X_0, X_2)^{-1}$  is given by

$$\frac{3}{16}\delta_{(-1,-4)} + \frac{1}{4}\delta_{(-1,0)} + \frac{1}{16}\delta_{(-1,4)} + \frac{1}{16}\delta_{(1,-4)} + \frac{1}{4}\delta_{(1,0)} + \frac{3}{16}\delta_{(1,4)}$$

whereas the unique left-monotone transport in  $\mathcal{M}(\mu_0, \mu_2)$  can be found to be

$$\frac{1}{8}\delta_{(-1,-4)} + \frac{3}{8}\delta_{(-1,0)} + \frac{1}{8}\delta_{(1,-4)} + \frac{1}{8}\delta_{(1,0)} + \frac{1}{4}\delta_{(1,4)}.$$

Therefore, there exists no transport  $P \in \mathcal{M}(\mu)$  such that both  $P_{01}$  and  $P_{02}$  are left-monotone, and Proposition 6.9 shows that (6.1) fails.

**Remark 6.11.** Of course, all our results on left-monotone transports have “right-monotone” analogues, obtained by reversing the orientation on the real line.

## 7 Geometry and Optimality Properties

In this section we introduce the optimality properties for transports and the geometric properties of their supports that were announced in the Introduction, and prove that they equivalently characterize left-monotone transports.

## 7.1 Geometry of Optimal Transports for Reward Functions of Spence–Mirrlees Type

The first goal is to show that optimal transports for specific reward functions are concentrated on sets  $\Gamma \subseteq \mathbb{R}^{n+1}$  satisfying certain no-crossing conditions that we introduce next. Given  $1 \leq t \leq n$ , we write

$$\Gamma^t = \{(x_0, \dots, x_t) \in \mathbb{R}^{t+1} : (x_0, \dots, x_n) \in \Gamma \text{ for some } (x_{t+1}, \dots, x_n) \in \mathbb{R}^{n-t}\}$$

for the projection of  $\Gamma$  onto the first  $t + 1$  coordinates.

**Definition 7.1.** Let  $\Gamma \subseteq \mathbb{R}^{n+1}$  and  $1 \leq t \leq n$ . Consider  $\mathbf{x} = (x_0, \dots, x_{t-1})$ ,  $\mathbf{x}' = (x'_0, \dots, x'_{t-1}) \in \mathbb{R}^t$  and  $y^+, y^-, y' \in \mathbb{R}$  with  $y^- < y^+$  such that  $(\mathbf{x}, y^+), (\mathbf{x}, y^-), (\mathbf{x}', y') \in \Gamma^t$ . Then, the projection

$$\Gamma^t \text{ is left-monotone if } y' \notin (y^-, y^+) \text{ whenever } x_0 < x'_0.$$

The set  $\Gamma$  is left-monotone<sup>6</sup> if  $\Gamma^t$  is left-monotone for all  $1 \leq t \leq n$ .

We also need the following notion.

**Definition 7.2.** Let  $\Gamma \subseteq \mathbb{R}^{n+1}$  and  $1 \leq t \leq n$ . The projection  $\Gamma^t$  is *nondegenerate* if for all  $\mathbf{x} = (x_0, \dots, x_{t-1}) \in \mathbb{R}^t$  and  $y \in \mathbb{R}$  such that  $(\mathbf{x}, y) \in \Gamma^t$ , the following hold:

- (i) if  $y > x_{t-1}$ , there exists  $y' < x_{t-1}$  such that  $(\mathbf{x}, y') \in \Gamma^t$ ;
- (ii) if  $y < x_{t-1}$ , there exists  $y' > x_{t-1}$  such that  $(\mathbf{x}, y') \in \Gamma^t$ .

The set  $\Gamma$  is called nondegenerate<sup>7</sup> if  $\Gamma^t$  is nondegenerate for all  $1 \leq t \leq n$ .

Broadly speaking, this definition says that for any up-path in  $\Gamma$  there exists a down-path, and vice versa. For a set supporting a martingale, nondegeneracy is not a restriction, in the following sense.

**Remark 7.3.** Let  $\Gamma \subseteq \mathbb{R}^{n+1}$  be a Borel set and  $\boldsymbol{\mu}$  in convex order.

(i) There exists a nondegenerate, universally measurable set  $\Gamma' \subseteq \Gamma$  such that  $P(\Gamma') = 1$  for all  $P \in \mathcal{M}(\boldsymbol{\mu})$  with  $P(\Gamma) = 1$ .

(ii) Fix  $P \in \mathcal{M}(\boldsymbol{\mu})$  with  $P(\Gamma) = 1$ . There exists a nondegenerate, Borel-measurable set  $\Gamma'_P \subseteq \Gamma$  such that  $P(\Gamma') = 1$ .

<sup>6</sup>This terminology is abusive since  $\Gamma = \Gamma^n$  is in fact a projection itself—it will be clear from the context what is meant.

<sup>7</sup>Footnote 6 applies here as well.

*Proof.* Let  $N_t$  be the set of all  $\mathbf{x} \in \Gamma^{t-1}$  such that (i) or (ii) of Definition 7.2 fail. If  $P$  is a martingale with  $P(\Gamma) = 1$ , we see that  $N_t \times \mathbb{R}^{n-t+1}$  is  $P$ -null. Moreover,  $N_t$  is universally measurable and we can set

$$\Gamma' := \Gamma \setminus \bigcup_{t=1}^n (N_t \times \mathbb{R}^{n-t+1})$$

to prove (i). Turning to (ii), universal measurability implies that there exists a Borel set  $N'_t \supseteq N_t$  such that  $N'_t \setminus N_t$  is  $P_{t-1}$ -null, where  $P_{t-1} = P \circ (X_0, \dots, X_{t-1})^{-1}$ . We can then set  $\Gamma'_P := \Gamma \setminus \bigcup_{t=1}^n (N'_t \times \mathbb{R}^{n-t+1})$ .  $\square$

Next, we introduce a notion of competitors along the lines of [8, Definition 1.10].

**Definition 7.4.** Let  $\pi$  be a finite measure on  $\mathbb{R}^{t+1}$  whose marginals have finite first moments and consider a disintegration  $\pi = \pi_t \otimes \kappa$ , where  $\pi_t$  is the projection of  $\pi$  onto the first  $t$  coordinates. A measure  $\pi' = \pi_t \otimes \kappa'$  is a  $t$ -competitor of  $\pi$  if it has the same last marginal and

$$\text{bary}(\kappa(\mathbf{x}, \cdot)) = \text{bary}(\kappa'(\mathbf{x}, \cdot)) \quad \text{for } \pi_t\text{-a.e. } \mathbf{x} = (x_0, \dots, x_{t-1}).$$

Using these definitions, we now formulate a variant of the monotonicity principle stated in Theorem 5.4 that will convenient to infer the geometry of  $\Gamma$ .

**Lemma 7.5.** Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order,  $1 \leq t \leq n$  and let  $\bar{f} : \mathbb{R}^{t+1} \rightarrow [0, \infty)$  be Borel. Consider  $f(X_0, \dots, X_n) := \bar{f}(X_0, \dots, X_t)$  and suppose that  $\mathbf{I}_{\boldsymbol{\mu}}(f) < \infty$ . Let  $(\phi, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$  be an optimizer for  $\mathbf{I}_{\boldsymbol{\mu}}(f)$  with the property that  $\phi_s \equiv H_s \equiv 0$  for  $s = t+1, \dots, n$  and define the set

$$\Gamma := \left\{ \sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n = f \right\} \cap \mathcal{V}.$$

Let  $\pi$  be a finitely supported probability on  $\mathbb{R}^{t+1}$  which is concentrated on  $\Gamma^t$ . Then  $\pi(\bar{f}) \geq \pi'(\bar{f})$  for any  $t$ -competitor  $\pi'$  of  $\pi$  that is concentrated on  $\mathcal{V}^t$ .

*Proof.* Recall that the projections  $\pi_t$  and  $\pi'_t$  onto the first  $t$  coordinates coincide. Thus,

$$\begin{aligned} \pi[H_t \cdot (X_t - X_{t-1})] &= \int H_t \cdot (\text{bary}(\kappa(X_0, \dots, X_{t-1}, \cdot) - X_{t-1}) d\pi_t \\ &= \int H_t \cdot (\text{bary}(\kappa'(X_0, \dots, X_{t-1}, \cdot) - X_{t-1}) d\pi'_t \\ &= \pi'[H_t \cdot (X_t - X_{t-1})]. \end{aligned}$$



Using also that the last marginals coincide, we deduce that

$$\pi[\bar{f}] = \pi \left[ \sum_{s=0}^t \phi_s(X_s) + (H \cdot X)_t \right] = \pi' \left[ \sum_{s=0}^t \phi_s(X_s) + (H \cdot X)_t \right] \geq \pi'[\bar{f}].$$

□

Next, we formulate an intermediate result relating optimality for Spence–Mirrlees reward functions to left-monotonicity of the support.

**Lemma 7.6.** *Let  $1 \leq t \leq n$  and let  $\Gamma \subseteq \mathcal{V}$  be a subset such that  $\Gamma^t$  is nondegenerate. Moreover, let  $f : \mathbb{R}^{t+1} \rightarrow \mathbb{R}$  be of the form  $f(X_0, \dots, X_t) = \bar{f}(X_0, X_t)$  for a second-order Spence–Mirrlees function  $\bar{f}$ . Assume that for any finitely supported probability  $\pi$  that is concentrated on  $\Gamma^t$  and any  $t$ -competitor  $\pi'$  of  $\pi$  that is concentrated on  $\mathcal{V}^t$ , we have  $\pi(f) \geq \pi'(f)$ . Then, the projection  $\Gamma^t$  is left-monotone.*

*Proof.* Consider  $(\mathbf{x}, y_1), (\mathbf{x}, y_2), (\mathbf{x}', y') \in \Gamma^t$  satisfying  $x_0 < x'_0$  and suppose for contradiction that  $y_1 < y' < y_2$ . We define  $\lambda = \frac{y_2 - y'}{y_2 - y_1}$  and

$$\begin{aligned} \pi &= \frac{\lambda}{2} \delta_{(\mathbf{x}, y_1)} + \frac{1 - \lambda}{2} \delta_{(\mathbf{x}, y_2)} + \frac{1}{2} \delta_{(\mathbf{x}', y')} \\ \pi' &= \frac{\lambda}{2} \delta_{(\mathbf{x}', y_1)} + \frac{1 - \lambda}{2} \delta_{(\mathbf{x}', y_2)} + \frac{1}{2} \delta_{(\mathbf{x}, y')}. \end{aligned}$$

Then  $\pi$  and  $\pi'$  have the same projection  $\pi_t = \pi'_t$  on the first  $t$  marginals and their last marginals also coincide. Moreover, disintegrating  $\pi = \pi_t \otimes \kappa$  and  $\pi' = \pi_t \otimes \kappa'$ , the measures  $\kappa(\mathbf{x}), \kappa(\mathbf{x}'), \kappa'(\mathbf{x}), \kappa'(\mathbf{x}')$  all have barycenter  $y'$ . Therefore,  $\pi$  and  $\pi'$  are  $t$ -competitors. We must also have that  $\pi'$  is concentrated on  $\mathcal{V}^t$ , by the shape of  $\mathcal{V}$ . Now our assumption implies that  $\pi(f) \geq \pi'(f)$ , but the second-order Spence–Mirrlees property of  $\bar{f}$  implies that  $\pi(f) < \pi'(f)$ . □

## 7.2 Geometry of Left-Monotone Transports

Next, we establish that transports with left-monotone support are indeed left-monotone in the sense of Theorem 6.8.

**Theorem 7.7.** *Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order and let  $P \in \mathcal{M}(\boldsymbol{\mu})$  be concentrated on a nondegenerate, left-monotone set  $\Gamma \subseteq \mathbb{R}^{n+1}$ . Then  $P$  is left-monotone.*

Before stating the proof of the theorem, we record two auxiliary results about measures on the real line. The first one is a direct consequence of Proposition 2.1.

**Lemma 7.8.** *Let  $a < b$  and  $\mu \leq_c \nu$ . If  $\nu$  is concentrated on  $(-\infty, a]$ , then so is  $\mu$ , and moreover  $\nu(\{a\}) \geq \mu(\{a\})$ . The analogue holds for  $[b, \infty)$ .*

The second result is [8, Lemma 5.2].

**Lemma 7.9.** *Let  $\sigma$  be a nontrivial signed measure on  $\mathbb{R}$  with  $\sigma(\mathbb{R}) = 0$  and let  $\sigma = \sigma^+ - \sigma^-$  be its Hahn decomposition. There exist  $a \in \text{supp}(\sigma^+)$  and  $b > a$  such that  $\int (b - y)^+ \mathbf{1}_{[a, \infty)} d\sigma(y) > 0$ .*

We can now give the proof of the theorem.

*Proof of Theorem 7.7.* Since the case  $n = 1$  is covered by Proposition 6.3, we may assume that the theorem has been proved for transports with  $n - 1$  steps and focus on the induction argument.

For every  $x \in \mathbb{R}$  we denote by  $\mu_x^t$  the marginal  $(P|_{(-\infty, x] \times \mathbb{R}^n}) \circ X_t^{-1}$ . In particular, we then have  $\mu_x^0 = \mu_0|_{(-\infty, x]}$  and  $\mu_x^t$  is the image of  $\mu_x^0$  under  $P$  after  $t$  steps. For the sake of brevity, we also set  $\nu_x^t := \mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_x^0)$ . By definition,  $P$  is left-monotone if  $\mu_x^t = \nu_x^t$  for all  $x \in \mathbb{R}$  and  $t \leq n$ , and by the induction hypothesis, we may assume that this holds for  $t \leq n - 1$ .

We argue by contradiction and assume that there exists  $x \in \mathbb{R}$  such that  $\mu_x^n \neq \nu_x^n$ . Then, the signed measure

$$\sigma := \nu_x^n - \mu_x^n$$

is nontrivial and we can find  $a < b$  with  $a \in \text{supp}(\sigma^+)$  as in Lemma 7.9. Observe that  $\sigma^+ \leq \mu_n - \mu_x^n$  where  $\mu_n - \mu_x^n$  is the image of  $\mu_n|_{(x, \infty)}$  under  $P$ . Hence,  $a \in \text{supp}(\mu_n - \mu_x^n)$  and as  $P$  is concentrated on  $\Gamma$ , we conclude that there exists a sequence of points

$$\mathbf{x}^m = (x_0^m, \dots, x_n^m) \in \Gamma \quad \text{with } x < x_0^m \text{ and } x_n^m \rightarrow a. \quad (7.1)$$

Moreover, by the characterization of the obstructed shadow in Lemma 6.7, we must have

$$\nu_x^n \leq_c \mu_x^n$$

as  $\mu_x^n \in [\mu_x^0; \mu_n]^{\mu_1, \dots, \mu_{n-1}}$  due to the fact that  $\mu_x^n$  is the image of  $\mu_x^0$  under a martingale transport.

*Step 1.* We claim that for all  $\mathbf{x} = (x_0, \dots, x_{n-1})$  with  $x_0 \leq x$  and  $x_{n-1} \leq a$ , it holds that

$$\Gamma_{\mathbf{x}} \cap (a, \infty) = \emptyset,$$

where  $\Gamma_{\mathbf{x}} = \{y \in \mathbb{R} : (\mathbf{x}, y) \in \Gamma\}$  is the section of  $\Gamma$  at  $\mathbf{x}$ . By way of contradiction, assume that for some  $\mathbf{x}$  with  $x_0 \leq x$  and  $x_{n-1} \leq a$  we have  $\Gamma_{\mathbf{x}} \cap (a, \infty) \neq \emptyset$ , then in particular  $\Gamma_{\mathbf{x}} \cap (x_{n-1}, \infty) \neq \emptyset$ . In view of the nondegeneracy of  $\Gamma$ , we conclude that  $\Gamma_{\mathbf{x}} \cap (-\infty, x_{n-1}) \neq \emptyset$  and hence that  $\Gamma_{\mathbf{x}} \cap (-\infty, a) \neq \emptyset$ . This yields a contradiction to the left-monotonicity of  $\Gamma$  by using  $\mathbf{x}^m$  from (7.1) for  $\mathbf{x}'$  in Definition 7.1 for large enough  $m$ , and the proof of the claim is complete.

*Step 2.* Similarly, we can show that for all  $\mathbf{x} = (x_0, \dots, x_{n-1})$  with  $x_0 \leq x$  and  $x_{n-1} \geq a$ ,

$$\Gamma_{\mathbf{x}} \cap (-\infty, a) = \emptyset.$$

*Step 3.* Next, we consider the marginals

$$\mu_{x,a}^t := (P|_{(-\infty, x] \times \mathbb{R}^{n-2} \times (-\infty, a] \times \mathbb{R}}) \circ X_t^{-1}.$$

Then, in particular,  $\mu_{x,a}^{n-1} = \mu_x^{n-1}|_{(-\infty, a]}$  and  $\mu_{x,a}^n$  is the image of  $\mu_{x,a}^{n-1}$  under the last step of  $P$ . Step 1 of the proof thus implies that  $\mu_{x,a}^n$  is concentrated on  $(-\infty, a]$ . We also write

$$\nu_{x,a}^n := \mathcal{S}^{\mu_x^n}(\mu_x^{n-1}|_{(-\infty, a]}).$$

We have  $\mu_{x,a}^{n-1} \leq_c \mu_{x,a}^n$  as  $\mathcal{M}(\mu_{x,a}^{n-1}, \mu_{x,a}^n) \neq \emptyset$ , and  $\mu_{x,a}^n \leq \mu_x^n$ . Therefore,

$$\nu_{x,a}^n \leq_c \mu_{x,a}^n \tag{7.2}$$

by the minimality of the shadow. Next, we show that

$$\nu_x^n - \nu_{x,a}^n \leq_c \mu_x^n - \mu_{x,a}^n. \tag{7.3}$$

Observe that  $\mu_x^n - \mu_{x,a}^n$  is the image of  $\mu_x^{n-1}|_{(a, \infty)}$  under  $P$  and therefore concentrated on  $[a, \infty)$  by Step 2. Using this observation, that  $\mu_{x,a}^n$  is concentrated on  $(-\infty, a]$  as mentioned above, and the fact that  $\nu_{x,a}^n(\{a\}) \leq \mu_{x,a}^n(\{a\})$  as a consequence of (7.2) and Lemma 7.8, we have

$$\mu_x^n - \mu_{x,a}^n = (\mu_x^n - \mu_{x,a}^n)|_{[a, \infty)} \leq (\mu_x^n - \mu_{x,a}^n)|_{[a, \infty)} \leq (\mu_x^n - \nu_{x,a}^n)|_{[a, \infty)} \leq \mu_x^n - \nu_{x,a}^n.$$

We also have  $\mu_x^{n-1}|_{(a, \infty)} \leq_c \mu_x^n - \mu_{x,a}^n$  since the latter measure is the image of the former under  $P$ . Together with the preceding display, we have established that

$$\mu_x^n - \mu_{x,a}^n \in \llbracket \mu_x^{n-1}|_{(a, \infty)}; \mu_x^n - \nu_{x,a}^n \rrbracket.$$

On the other hand,

$$\nu_x^n - \nu_{x,a}^n = \mathcal{S}^{\mu_x^n - \nu_{x,a}^n}(\mu_x^{n-1}|_{(a, \infty)})$$

from the additivity property of the shadow in Lemma 6.5(i), and therefore (7.3) follows by the minimality of the shadow.

*Step 4.* Recall from Step 3 that  $\mu_{x,a}^n$  is concentrated on  $(-\infty, a]$  and that  $\mu_x^n - \mu_{x,a}^n$  is concentrated on  $[a, \infty)$ . Therefore,  $\nu_{x,a}^n$  is concentrated on  $(-\infty, a]$  and  $\nu_x^n - \nu_{x,a}^n$  is concentrated on  $[a, \infty)$ , by Lemma 7.8. Moreover, we have  $\nu_{x,a}^n(\{a\}) \leq \mu_{x,a}^n(\{a\})$  by the same lemma, and finally, the function  $y \mapsto (b-y)^+ \mathbf{1}_{[a,\infty)}(y)$  is convex on  $[a, \infty)$  as  $a < b$ . Using these facts and (7.3),

$$\begin{aligned} & \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y) \nu_x^n(dy) \\ &= \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y) (\nu_x^n - \nu_{x,a}^n)(dy) + (b-a) \nu_{x,a}^n(\{a\}) \\ &\leq \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y) (\mu_x^n - \mu_{x,a}^n)(dy) + (b-a) \mu_{x,a}^n(\{a\}) \\ &= \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y) \mu_x^n(dy). \end{aligned}$$

This contradicts the choice of  $a$  and  $b$ , cf. Lemma 7.9, and thus completes the proof.  $\square$

### 7.3 Optimality Properties

In this section we relate left-monotone transports and left-monotone sets to the optimal transport problem for Spence–Mirrlees functions.

**Theorem 7.10.** *For  $1 \leq t \leq n$ , let  $f_t : \mathbb{R}^2 \rightarrow \mathbb{R}$  be second-order Spence–Mirrlees functions such that  $|f_t(x, y)| \leq a_0(x) + a_t(y)$  for some  $a_0 \in L^1(\mu_0)$  and  $a_t \in L^1(\mu_t)$ . There exists a universally measurable, nondegenerate, left-monotone set  $\Gamma' \subseteq \mathbb{R}^{n+1}$  such that any simultaneous optimizer  $P \in \mathcal{M}(\boldsymbol{\mu})$  for  $\mathbf{S}_{\boldsymbol{\mu}}(f_t(X_0, X_t))$ ,  $1 \leq t \leq n$  is concentrated on  $\Gamma'$ . In particular, any such  $P$  is left-monotone.*

*Proof.* The last assertion follows by an application of Theorem 7.7, so we may focus on finding  $\Gamma'$ . For each  $1 \leq t \leq n$ , we use Theorem 5.2 and Remark 5.3 to find a dual optimizer  $(\phi, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$  for  $\mathbf{I}_{\boldsymbol{\mu}}(f_t(X_0, X_t))$  and define the Borel set

$$\Gamma(t) := \left\{ \sum_{s=0}^n \phi_s(X_s) + (H \cdot X)_n = f \right\} \cap \mathcal{V}.$$

Here, we may choose a dual optimizer such that  $\phi_s \equiv H_s \equiv 0$  for  $s = t+1, \dots, n$ . (This can be seen by applying Theorem 5.2 to the transport

problem involving only the marginals  $(\mu_0, \dots, \mu_t)$  and taking the corresponding dual optimizer.) Theorem 5.4 shows that any simultaneous optimizer  $P \in \mathcal{M}(\boldsymbol{\mu})$  is concentrated on  $\Gamma(t)$  for all  $t$ , and hence also on the Borel set

$$\Gamma := \bigcap_{t=1}^n \Gamma(t).$$

Using Remark 7.3 (i), we find a universally measurable, nondegenerate subset  $\Gamma' \subseteq \Gamma$  with the same property. Since the projection  $(\Gamma')^t$  is contained in the projection  $(\Gamma(t))^t$ , Lemma 7.5 and Lemma 7.6 yield that  $(\Gamma')^t$  is left-monotone for all  $t$ ; that is,  $\Gamma'$  is left-monotone.  $\square$

**Remark 7.11.** In Theorem 7.10, if we only wish to find a nondegenerate, left-monotone set  $\Gamma'_P \subseteq \mathbb{R}^{n+1}$  such that a given simultaneous optimizer  $P \in \mathcal{M}(\boldsymbol{\mu})$  is concentrated on  $\Gamma'_P$ , then we may choose  $\Gamma'_P$  to be Borel instead of universally measurable. This follows by replacing the application of Remark 7.3 (i) by Remark 7.3 (ii) in the proof.

The following is a converse to Theorem 7.10.

**Theorem 7.12.** *Given  $1 \leq t \leq n$ , let  $f \in C^{1,2}(\mathbb{R}^2)$  be such that  $f_{xyy} \geq 0$  and suppose that the following integrability condition holds:*

$$\begin{cases} f(X_0, X_t), & f(0, X_t), & f(X_0, 0), & \bar{h}(X_0)X_0, & \bar{h}(X_0)X_t \\ \text{are } P\text{-integrable for all } P \in \mathcal{M}(\boldsymbol{\mu}), \end{cases} \quad (7.4)$$

where  $\bar{h}(x) := \partial_y|_{y=0}[f(x, y) - f(0, y)]$ . Then every left-monotone transport  $P \in \mathcal{M}(\boldsymbol{\mu})$  is an optimizer for  $\mathbf{S}_{\boldsymbol{\mu}}(f)$ .

The integrability condition clearly holds when  $f$  is Lipschitz continuous; in particular, a smooth second-order Spence–Mirrlees function (as defined in the Introduction) satisfies the assumptions of the theorem for any  $\boldsymbol{\mu}$ .

The proof will be given by an approximation based on the following building blocks for Spence–Mirrlees functions; the construction is novel and may be of independent interest.

**Lemma 7.13.** *Let  $1 \leq t \leq n$  and let  $f(X_0, \dots, X_n) := \mathbf{1}_{(-\infty, a]}(X_0)\varphi(X_t)$  for a concave function  $\varphi$  and  $a \in \mathbb{R}$ . Then every left-monotone transport  $P \in \mathcal{M}(\boldsymbol{\mu})$  is an optimizer for  $\mathbf{S}_{\boldsymbol{\mu}}(f)$ .*

*Proof.* In view of Lemma 6.7, this follows directly by applying the defining shadow property from Theorem 6.8 with  $x = a$ .  $\square$

The integrability condition (7.4) implies that setting

$$g(x, y) := f(x, 0) + f(0, y) - f(0, 0) + \bar{h}(x)y,$$

the three terms constituting

$$g(X_0, X_t) = [f(X_0, 0) + \bar{h}(X_0)X_0] + [f(0, X_t) - f(0, 0)] + [\bar{h}(X_0)(X_t - X_0)]$$

are  $P$ -integrable and  $P[g(X_0, X_t)]$  is constant over  $P \in \mathcal{M}(\boldsymbol{\mu})$ . By replacing  $f$  with  $f - g$ , we may thus assume without loss of generality that

$$f(x, 0) = f(0, y) = f_y(x, 0) = 0 \quad \text{for all } (x, y) \in \mathbb{R}^2. \quad (7.5)$$

After this normalization, integration by parts yields the representation

$$f(x, y) = \int_0^y \int_0^x (y - t) f_{xyy}(s, t) ds dt. \quad (7.6)$$

**Lemma 7.14.** *Theorem 7.12 holds under the following additional condition: there exists a constant  $c > 0$  such that*

$$\begin{aligned} x \mapsto f(x, y) &\text{ is constant on } \{x > c\} \text{ and on } \{x < -c\}, \\ y \mapsto f(x, y) &\text{ is affine on } \{y > c\} \text{ and on } \{y < -c\}. \end{aligned}$$

*Proof.* The additional condition and integration by parts imply that for all  $(x, y) \in \mathbb{R}^2$ , we have the representation

$$\begin{aligned} f(x, y) = & - \int_{-c}^c \int_{-c}^c \mathbf{1}_{(-\infty, s]}(x)(y - t)^+ f_{xyy}(s, t) ds dt \\ & + [f(x, -c) - (-c)f_y(x, -c)] \\ & + [f(c, y) - f(c, -c) - f_y(c, -c)(y - (-c))] \\ & + f_y(x, -c)y. \end{aligned}$$

The last three terms are of the form  $g(x, y) = \tilde{\phi}(x) + \tilde{\psi}(y) + \tilde{h}(x)y$  and of linear growth due to the additional condition. Hence, as above,  $P'[g(X_0, X_t)] = C$  is constant for  $P' \in \mathcal{M}(\boldsymbol{\mu})$ . If  $P \in \mathcal{M}(\boldsymbol{\mu})$  is left-monotone and  $P' \in \mathcal{M}(\boldsymbol{\mu})$  is arbitrary, Fubini's theorem and Lemma 7.13 yield that

$$\begin{aligned} P[f] &= - \int_{-c}^c \int_{-c}^c P[\mathbf{1}_{(-\infty, s]}(x)(y - t)^+] f_{xyy}(s, t) ds dt + C \\ &\geq - \int_{-c}^c \int_{-c}^c P'[\mathbf{1}_{(-\infty, s]}(x)(y - t)^+] f_{xyy}(s, t) ds dt + C \\ &= P'[f], \end{aligned}$$

where  $P, P'$  are understood to integrate with respect to  $(x, y)$  and the application of Fubini's theorem is justified by the nonnegativity of the integrand.  $\square$

*Proof of Theorem 7.12.* Let  $f$  be as in the theorem. We shall construct functions  $f^m$ ,  $m \geq 1$  satisfying the assumption of Lemma 7.14 as well as  $P[f^m] \rightarrow P[f]$  for all  $P \in \mathcal{M}(\boldsymbol{\mu})$ . Once this is achieved, the theorem follows from the lemma.

Indeed, we may assume that  $f$  is normalized as in (7.5). Let  $m \geq 1$  and let  $\rho_m : \mathbb{R} \rightarrow [0, 1]$  be a smooth function such that  $\rho_m = 1$  on  $[-m, m]$  and  $\rho_m = 0$  on  $[-m-1, m+1]^c$ . In view of (7.6), we define  $f^m$  by

$$f^m(x, y) = \int_0^y \int_0^x (y-t) f_{xyy}(s, t) \rho_m(s) \rho_m(t) ds dt.$$

It then follows that  $f^m$  satisfies the assumptions of Lemma 7.14 with the constant  $c = m+1$ . Moreover, we have

$$0 \leq f^m(x, y) \leq f^{m+1}(x, y) \leq f(x, y) \quad \text{for } x \geq 0$$

and the opposite inequalities for  $x \leq 0$ , as well as  $f^m(x, y) \rightarrow f(x, y)$  for all  $(x, y)$ .

Let  $P \in \mathcal{M}(\boldsymbol{\mu})$ . Since  $f$  is  $P$ -integrable, applying monotone convergence separately on  $\{x \geq 0\}$  and  $\{x \leq 0\}$  yields that  $P[f^m] \rightarrow P[f]$ , and the proof is complete.  $\square$

**Remark 7.15.** The function

$$\bar{f}(x, y) := \tanh(x) \sqrt{1+y^2}$$

satisfies the conditions of Theorem 7.12 for all marginals  $\boldsymbol{\mu}$  in convex order, since the latter are assumed to have a finite first moment.

We can now collect the preceding results to obtain, in particular, the equivalences stated in Theorem 1.1.

**Theorem 7.16.** *Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order. There exists a left-monotone, nondegenerate, universally measurable set  $\Gamma \subseteq \mathbb{R}^{n+1}$  such that for any  $P \in \mathcal{M}(\boldsymbol{\mu})$ , the following are equivalent:*

- (i)  *$P$  is an optimizer for  $\mathbf{S}_{\boldsymbol{\mu}}(f(X_0, X_t))$  whenever  $f$  is a smooth second-order Spence–Mirrlees function and  $1 \leq t \leq n$ ,*
- (ii)  *$P$  is concentrated on  $\Gamma$ ,*

(ii')  $P$  is concentrated on a left-monotone set,

(iii)  $P$  is left-monotone; i.e.  $P_{0t}$  transports  $\mu_0|_{(-\infty, a]}$  to  $\mathcal{S}^{\mu_1, \dots, \mu_t}(\mu_0|_{(-\infty, a]})$  for all  $1 \leq t \leq n$  and  $a \in \mathbb{R}$ .

Moreover, there exists  $P \in \mathcal{M}(\boldsymbol{\mu})$  satisfying (i)–(iii).

*Proof.* Let  $\Gamma$  be the set provided by Theorem 7.10 for the function  $f_t = \bar{f}$  of Remark 7.15. Given  $P \in \mathcal{M}(\boldsymbol{\mu})$ , Theorem 7.10 shows that (i) implies (ii) which trivially implies (ii'). Theorem 7.7 and Remark 7.3 show that (ii') implies (iii), and Theorem 7.12 shows that (iii) implies (i). Finally, the existence of a left-monotone transport was stated in Theorem 6.8.  $\square$

We conclude this section with an example showing that left-monotone transports are not Markovian in general, even if they are unique and (6.1) holds for  $\boldsymbol{\mu}$ .

**Example 7.17.** Consider the marginals

$$\mu_0 = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \quad \mu_1 = \frac{3}{4}\delta_0 + \frac{1}{4}\delta_2, \quad \mu_2 = \frac{1}{8}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{8}\delta_1 + \frac{1}{4}\delta_2.$$

The transport  $P \in \mathcal{M}(\boldsymbol{\mu})$  given by

$$P = \frac{1}{2}\delta_{(0,0,0)} + \frac{1}{8}\delta_{(1,0,-1)} + \frac{1}{8}\delta_{(1,0,1)} + \frac{1}{4}\delta_{(1,2,2)}$$

is left-monotone because its support is left-monotone (Figure 4), and it is clearly not Markovian. On the other hand, it is not hard to see that this is the only way to build a left-monotone transport in  $\mathcal{M}(\boldsymbol{\mu})$ .

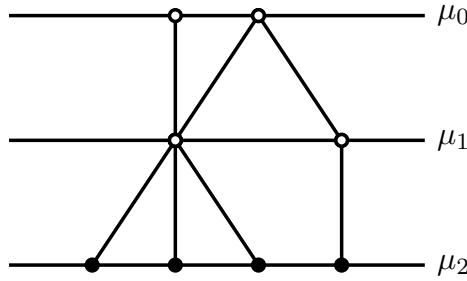


Figure 4: Support of the non-Markovian transport in Example 7.17.



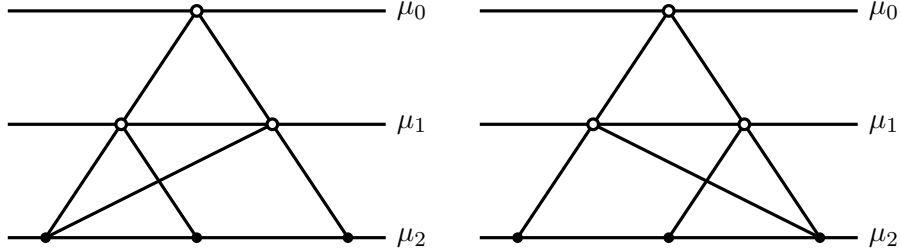


Figure 5: Supports of two left-monotone transports for the same marginals.

## 8 Uniqueness of Left-Monotone Transports

In this section we consider the (non-)uniqueness of left-monotone transports. It turns out the the presence of atoms in  $\mu_0$  is important in this respect—let us start with the following simple observation.

**Remark 8.1.** Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order. If  $\mu_0$  is a Dirac mass, then every  $P \in \mathcal{M}(\boldsymbol{\mu})$  is left-monotone. Indeed,  $\mathcal{M}(\mu_0, \mu_t)$  is a singleton for every  $1 \leq t \leq n$ , hence  $P_{0t}$  must be the (one-step) left-monotone transport.

Exploiting this observation, the following shows that left-monotone transports need not be unique when  $n \geq 2$ .

**Example 8.2.** Let  $\mu_0 = \delta_0$ ,  $\mu_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ ,  $\mu_2 = \frac{3}{8}\delta_{-2} + \frac{1}{4}\delta_0 + \frac{3}{8}\delta_2$ . By the remark, any element in  $\mathcal{M}(\boldsymbol{\mu})$  is left-monotone. Moreover,  $\mathcal{M}(\boldsymbol{\mu})$  is a continuum since  $\mathcal{M}(\mu_1, \mu_2)$  contains the convex hull of the two measures

$$\begin{aligned} P_l &= \frac{1}{4}\delta_{(-1,-2)} + \frac{1}{4}\delta_{(-1,0)} + \frac{1}{8}\delta_{(1,-2)} + \frac{3}{8}\delta_{(1,2)}, \\ P_r &= \frac{3}{8}\delta_{(-1,-2)} + \frac{1}{8}\delta_{(-1,2)} + \frac{1}{4}\delta_{(1,0)} + \frac{1}{4}\delta_{(1,2)}. \end{aligned}$$

The corresponding supports are depicted in Figure 5.

The example illustrates that non-uniqueness can typically be expected when  $\mu_0$  has atoms. On the other hand, we have the following uniqueness result.

**Theorem 8.3.** *Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order. If  $\mu_0$  is atomless, there exists a unique left-monotone transport  $P \in \mathcal{M}(\boldsymbol{\mu})$ .*

The remainder of this section is devoted to the proof. Let us call a kernel  $\kappa(x, dy)$  *binomial* if for all  $x \in \mathbb{R}$ , the measure  $\kappa(x, dy)$  consists of (at most) two point masses. A martingale transport will be called binomial if it can be disintegrated using only binomial kernels. We shall show that when  $\mu_0$  is atomless, any left-monotone transport is a binomial martingale, and then conclude the uniqueness via a convexity argument.

The first step is the following set-theoretic result.

**Lemma 8.4.** *Let  $k \geq 1$  be an integer and  $\Gamma \subseteq \mathbb{R}^{t+1}$ . For  $\mathbf{x} \in \mathbb{R}^t$ , we denote by  $\Gamma_{\mathbf{x}} := \{y \in \mathbb{R} : (\mathbf{x}, y) \in \Gamma\}$  the section at  $\mathbf{x}$ . If the set*

$$\{\mathbf{x} \in \mathbb{R}^t : |\Gamma_{\mathbf{x}}| \geq k\}$$

*is uncountable, then it has an accumulation point. More precisely, there are  $\mathbf{x} = (x_0, \dots, x_t) \in \mathbb{R}^t$  and  $y_1 < \dots < y_k$  in  $\Gamma_{\mathbf{x}}$  such that for all  $\epsilon > 0$  there exist  $\mathbf{x}' = (x'_0, \dots, x'_t) \in \mathbb{R}^t$  and  $y'_1 < \dots < y'_k$  in  $\Gamma_{\mathbf{x}'}$  satisfying*

$$(i) \quad \|\mathbf{x} - \mathbf{x}'\| < \epsilon,$$

$$(ii) \quad x_0 < x'_0,$$

$$(iii) \quad \max_{i=1, \dots, k} |y_i - y'_i| < \epsilon.$$

*Proof.* The proof is similar to the one of [8, Lemma 3.2] and therefore omitted.  $\square$

The following statement on the binomial structure generalizes a result of [8] for the one-step case and is of independent interest.

**Proposition 8.5.** *Let  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  be in convex order and let  $\mu_0$  be atomless. There exists a universally measurable set  $\Gamma \subseteq \mathbb{R}^{n+1}$  such that every left-monotone transport  $P \in \mathcal{M}(\boldsymbol{\mu})$  is concentrated on  $\Gamma$  and such that for all  $1 \leq t \leq n$  and  $\mathbf{x} \in \mathbb{R}^t$ ,*

$$|\{y \in \mathbb{R} : (X_0, \dots, X_t)^{-1}(\mathbf{x}, y) \cap \Gamma \neq \emptyset\}| \leq 2. \quad (8.1)$$

*In particular, every left-monotone transport  $P \in \mathcal{M}(\boldsymbol{\mu})$  is a binomial martingale.*

*Proof.* Let  $\Gamma$  be as in Theorem 7.16; then every left-monotone  $P \in \mathcal{M}(\boldsymbol{\mu})$  is concentrated on  $\Gamma$ . Let  $A_t$  be the set of all  $\mathbf{x} \in \mathbb{R}^t$  such that (8.1) fails. Suppose that  $A_t$  is uncountable; then Lemma 8.4 yields points  $\mathbf{x}, \mathbf{x}'$  such that for some  $y_1, y_2 \in \Gamma_{\mathbf{x}}^t$  and  $y \in \Gamma_{\mathbf{x}'}^t$ , we have  $y_1 < y < y_2$ . This contradicts the left-monotonicity of  $\Gamma$  (Definition 7.1), thus  $A_t$  must be countable. Hence,

$(X_0, \dots, X_{t-1})^{-1}(A_t)$  is Borel and  $P$ -null for all  $P \in \mathcal{M}(\boldsymbol{\mu})$ , as  $\mu_0$  is atomless. The set  $\Gamma' = \Gamma \setminus \cup_{t=1}^n (X_0, \dots, X_{t-1})^{-1}(A_t)$  then has the required properties.  $\square$

*Proof of Theorem 8.3.* We will prove this result using induction on  $n$ . For  $n = 1$  the result holds by Proposition 6.3, with or without atoms. To show the induction step, let  $P'$  be the unique left-monotone transport in  $\mathcal{M}(\mu_0, \dots, \mu_{n-1})$  and let  $P_1 = P' \otimes \kappa_1$  and  $P_2 = P' \otimes \kappa_2$  be disintegrations of two  $n$ -step left-monotone transports. Then,

$$\frac{P_1 + P_2}{2} = P' \otimes \frac{\kappa_1 + \kappa_2}{2}$$

is again left-monotone, and Proposition 8.5 yields that  $(\kappa_1 + \kappa_2)/2$  must be a binomial kernel  $P'$ -a.s. Using also the martingale property of  $\kappa_1$  and  $\kappa_2$ , this can only be true if  $\kappa_1 = \kappa_2$  holds  $P'$ -a.s., and therefore  $P_1 = P_2$ .  $\square$

## 9 Free Intermediate Marginals

In this section we discuss a variant of our transport problem where the intermediate marginal constraints  $\mu_1, \dots, \mu_{n-1}$  are omitted; that is, only the first and last marginals  $\mu_0, \mu_n$  are prescribed. (One could similarly adapt the results to a case where some, but not all of the intermediate marginals are given.)

The primal space will be denoted by  $\mathcal{M}^n(\mu_0, \mu_n)$  and consists of all martingale measures  $P$  on  $\mathbb{R}^{n+1}$  such that  $\mu_0 = P \circ (X_0)^{-1}$  and  $\mu_n = P \circ (X_n)^{-1}$ . To make the connection with the previous sections, we note that

$$\mathcal{M}^n(\mu_0, \mu_n) = \bigcup \mathcal{M}(\boldsymbol{\mu})$$

where the union is taken over all vectors  $\boldsymbol{\mu} = (\mu_0, \mu_1, \dots, \mu_{n-1}, \mu_n)$  in convex order.

### 9.1 Polar Structure

We first characterize the polar sets of  $\mathcal{M}^n(\mu_0, \mu_n)$ . To that end, we introduce an analogue of the irreducible components.

**Definition 9.1.** Let  $\mu_0 \leq_c \mu_n$  and let  $(I_k, J_k) \subseteq \mathbb{R}^2$  be the corresponding irreducible domains in the sense of Proposition 2.3. The  $n$ -step components of  $\mathcal{M}^n(\mu_0, \mu_n)$  are the sets<sup>8</sup>

---

<sup>8</sup>A superscript  $m$  indicates the  $m$ -fold Cartesian product;  $\Delta_n$  is the diagonal in  $\mathbb{R}^{n+1}$ .

- (i)  $I_k^n \times J_k$ , where  $k \geq 1$ ,
- (ii)  $I_0^{n+1} \cap \Delta_n$ ,
- (iii)  $I_k^t \times \{p\}^{n-t+1}$ , where  $p \in J_k \setminus I_k$  and  $1 \leq t \leq n$ ,  $k \geq 1$ .

The characterization then takes the following form.

**Theorem 9.2** (Polar Structure). *Let  $\mu_0 \leq_c \mu_n$ . A Borel set  $B \subseteq \mathbb{R}^{n+1}$  is  $\mathcal{M}^n(\mu_0, \mu_n)$ -polar if and only if there exist a  $\mu_0$ -nullset  $N_0$  and a  $\mu_n$ -nullset  $N_n$  such that*

$$B \subseteq (N_0 \times \mathbb{R}^n) \cup (\mathbb{R}^n \times N_n) \cup \left( \bigcup V_j \right)^c$$

where the union runs over all  $n$ -step components  $V_j$  of  $\mathcal{M}^n(\mu_0, \mu_n)$ .

It turns out that our previous results can be put to work to prove the theorem, by means of the following lemma which may be of independent interest.

**Lemma 9.3.** *Let  $\mu \leq_c \nu$  be irreducible with domain  $(I, J)$  and let  $\rho$  be a probability concentrated on  $J$ . Then, there exists a probability  $\mu \leq_c \theta \leq_c \nu$  satisfying  $\theta \gg \rho$  such that  $\mu \leq_c \theta$  and  $\theta|_I \leq_c (\nu - \theta|_{J \setminus I})$  are both irreducible.*

*Proof. Step 1.* We first assume that  $\rho = \delta_x$  for some  $x \in J$  and show that there exists  $\theta$  satisfying

$$\mu \leq_c \theta \leq_c \nu \quad \text{and} \quad \theta \gg \delta_x.$$

If  $\nu$  has an atom at  $x$ , we can choose  $\theta = \nu$ . Thus, we may assume that  $\nu(\{x\}) = 0$  and in particular that  $x \in I$ . Let  $a$  be the common barycenter of  $\mu$  and  $\nu$  and suppose that  $x < a$ . For all  $b \in \mathbb{R}$  and  $0 \leq c \leq \nu(\{b\})$ , the measure

$$\nu_{b,c} := \nu|_{(-\infty, b)} + c\delta_b$$

satisfies  $\nu_{b,c} \leq \nu$ , and as  $x < a$  there are unique  $b, c$  such that  $\text{bary}(\nu_{b,c}) = x$ . Setting  $\alpha = \nu_{b,c}$  and  $\epsilon_0 = \alpha(\mathbb{R})$ , we then have  $\epsilon_0 \delta_x \leq_c \alpha \leq \nu$ , and a similar construction yields this result for  $x \geq a$ . The existence of such  $\alpha$  implies that

$$\epsilon \delta_x \leq_{pc} \nu, \quad 0 \leq \epsilon \leq \epsilon_0$$

and thus the shadow  $\mathcal{S}^\nu(\epsilon \delta_x)$  is well-defined. This measure is given by the restriction of  $\nu$  to an interval (possibly including fractions of atoms at the endpoints); cf. [8, Example 4.7]. Moreover, the interval is bounded after

possibly reducing the mass  $\epsilon_0$ . Thus, for all  $\epsilon < \epsilon_0$ , the difference of potential functions

$$u_{\mathcal{S}^\nu(\epsilon\delta_x)} - u_{\epsilon\delta_x} \geq 0$$

vanishes outside a compact interval, and it converges uniformly to zero as  $\epsilon \rightarrow 0$ .

On the other hand, as  $\mu \leq_c \nu$  is irreducible, the difference  $u_\nu - u_\mu \geq 0$  is uniformly bounded away from zero on compact subsets of  $I$  and has nonzero derivative on  $J \setminus I$ . Together, it follows that

$$u_\nu - u_{\mathcal{S}^\nu(\epsilon\delta_x)} + u_{\epsilon\delta_x} \geq u_\mu$$

for small enough  $\epsilon > 0$ , so that

$$\theta := \nu - \mathcal{S}^\nu(\epsilon\delta_x) + \epsilon\delta_x$$

satisfies  $\mu \leq_c \theta \leq_c \nu$ ; moreover,  $\theta \gg \delta_x$  as  $\nu(\{x\}) = 0$ .

*Step 2.* We turn to the case of a general probability measure  $\rho$  on  $J$ . By Step 1, we can find a measure  $\theta_x$  for each  $x \in J$  such that

$$\mu \leq_c \theta_x \leq_c \nu \quad \text{and} \quad \theta_x \gg \delta_x.$$

The map  $x \mapsto \theta_x$  can easily be chosen to be measurable (by choosing  $\epsilon$  in a measurable way). We can then define the probability measure

$$\theta'(A) := \int_J \theta_x(A) \rho(dx), \quad A \in \mathfrak{B}(\mathbb{R})$$

which satisfies  $\mu \leq_c \theta' \leq_c \nu$ . Moreover, we have  $\theta' \gg \delta_x$ ; indeed, if  $A \in \mathfrak{B}(\mathbb{R})$  is a  $\theta'$ -nullset, then  $\theta_x(A) = 0$  for  $\rho$ -a.e.  $x$  and thus  $\rho(A) = 0$  as  $\theta_x \gg \delta_x$ .

Finally,  $\theta := (\mu + \theta' + \nu)/3$  shares these properties. As  $u_\mu < u_\nu$  on  $I$  due to irreducibility, we have  $u_\mu < u_\theta < u_\nu$  on  $I$  and it follows that  $\mu \leq_c \theta$  and  $\theta|_I \leq_c (\nu - \theta|_{J \setminus I})$  are irreducible.  $\square$

**Lemma 9.4.** *Let  $\mu_0 \leq_c \mu_n$  and let  $\pi$  be a measure on  $\mathbb{R}^{n+1}$  which is concentrated on an  $n$ -step component  $V$  of  $\mathcal{M}^n(\mu_0, \mu_n)$  and whose first and last marginals satisfy*

$$\pi_0 \leq \mu_0, \quad \pi_n \leq \mu_n.$$

*Then there exists  $P \in \mathcal{M}^n(\mu_0, \mu_n)$  such that  $P \gg \pi$ .*

*Proof.* If  $V = I_0^{n+1} \cap \Delta_n$ , then  $\pi$  must be an identical transport and we can take  $P$  to be any element of  $\mathcal{M}(\mu_0, \mu_0, \dots, \mu_0, \mu_n)$ . Thus, we may assume

that  $V$  is of type (i) or (iii) in Definition 9.1, and then, by fixing  $k \geq 1$ , that  $\mu_0 \leq_c \mu_n$  is irreducible with domain  $(I, J)$ .

Using Lemma 9.3, we can find intermediate marginals  $\mu_t$  with

$$\mu_0 \leq_c \mu_1 \leq_c \cdots \leq_c \mu_{n-1} \leq \mu_n$$

such that  $\mu_t \gg \pi_t$  for all  $1 \leq t \leq n-1$ , and each of the steps  $\mu_{t-1} \leq_c \mu_t$ ,  $1 \leq t \leq n$  has a single irreducible domain given by  $(I, J)$  as well as (possibly) a diagonal component on  $J \setminus I$ . We note that  $V$  is an irreducible component of  $\mathcal{M}(\mu_0, \mu_1, \dots, \mu_n)$  as introduced after Theorem 3.1.

Let  $f_t = d\pi_t/d\mu_t$  be the Radon–Nikodym derivative of the marginal at date  $t$ . For  $m \geq 1$ , we define the measure  $\pi^m \ll \pi$  by

$$\pi^m(dx_0, \dots, dx_n) = 2^{-m} \left( \prod_{t=1}^{n-1} \mathbf{1}_{f_t(x_t) \leq 2^m} \right) \pi(dx_0, \dots, dx_n).$$

Then, the marginals  $\pi_t^m$  satisfy the stronger condition  $\pi_t^m \leq \mu_t$  for  $0 \leq t \leq n$ . Thus, we can apply Lemma 3.3 to  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  and the irreducible component  $V$ , to find  $P^m \in \mathcal{M}(\boldsymbol{\mu}) \subseteq \mathcal{M}^n(\mu_0, \mu_n)$  such that  $P^m \gg \pi^m$ . Noting that  $\sum_{m \geq 1} 2^{-m} \pi^m \gg \pi$ , we see that  $P := \sum_{m \geq 1} 2^{-m} P^m \gg \pi$  satisfies the requirements of the lemma.  $\square$

*Proof of Theorem 9.2.* The result is deduced from Lemma 9.4 by following the argument in the proof of Theorem 3.1.  $\square$

## 9.2 Duality

In this section we formulate a duality theorem for the transport problem with free intermediate marginals.

**Definition 9.5.** Let  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$ . The *primal problem* is

$$\mathbf{S}_{\mu_0, \mu_n}^n(f) := \sup_{P \in \mathcal{M}^n(\mu_0, \mu_n)} P(f) \in [0, \infty]$$

and the dual problem is

$$\mathbf{I}_{\mu_0, \mu_n}^n(f) := \inf_{(\phi, \psi, H) \in \mathcal{D}_{\mu_0, \mu_n}^n(f)} \mu_0(\phi) + \mu_n(\psi) \in [0, \infty],$$

where  $\mathcal{D}_{\mu_0, \mu_n}^n(f)$  consists of all triplets  $(\phi, \psi, H)$  such that  $(\phi, \psi) \in L^c(\mu_0, \mu_n)$  and  $H = (H_1, \dots, H_n)$  is  $\mathbb{F}$ -predictable with

$$\phi(X_0) + \psi(X_n) + (H \cdot X)_n \geq f \quad \mathcal{M}^n(\mu_0, \mu_n)\text{-q.s.}$$

i.e. the inequality holds  $P$ -a.s. for all  $P \in \mathcal{M}^n(\mu_0, \mu_n)$ .

The analogue of Theorem 5.2 reads as follows.

**Theorem 9.6** (Duality). *Let  $f : \mathbb{R}^{n+1} \rightarrow [0, \infty]$ .*

- (i) *If  $f$  is upper semianalytic, then  $\mathbf{S}_{\mu_0, \mu_n}^n(f) = \mathbf{I}_{\mu_0, \mu_n}^n(f) \in [0, \infty]$ .*
- (ii) *If  $\mathbf{I}_{\mu_0, \mu_n}^n(f) < \infty$ , there exists a dual optimizer  $(\phi, \psi, H) \in \mathcal{D}_{\mu_0, \mu_n}^n(f)$ .*

The main step for the proof is again a closedness result. We shall only discuss the case where  $\mu_0 \leq_c \mu_n$  is irreducible; the extension to the general case can be obtained along the lines of Section 4.

**Proposition 9.7.** *Let  $\mu_0 \leq_c \mu_n$  be irreducible and let  $f^m : \mathbb{R}^{n+1} \rightarrow [0, \infty]$  be a sequence of functions such that  $f^m \rightarrow f$  pointwise. Moreover, let  $(\phi^m, \psi^m, H^m) \in \mathcal{D}_{\mu_0, \mu_n}^n(f^m)$  be such that  $\sup_m \mu_0(\phi^m) + \mu_n(\psi^m) < \infty$ . Then there exist  $(\phi, \psi, H) \in \mathcal{D}_{\mu_0, \mu_n}^n(f)$  such that*

$$\mu_0(\phi) + \mu_n(\psi) \leq \liminf_{m \rightarrow \infty} \mu_0(\phi^m) + \mu_n(\psi^m).$$

*Proof.* Let  $\mu_t$ ,  $1 \leq t \leq n-1$  be such that  $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$  is in convex order and  $\mu_{t-1} \leq_c \mu_t$  is irreducible for all  $1 \leq t \leq n$ ; such  $\mu_t$  are easily constructed by prescribing their potential functions. Setting  $\boldsymbol{\phi}^m = (\phi^m, 0, \dots, 0, \psi^m)$  we have  $(\boldsymbol{\phi}^m, H^m) \in \mathcal{D}_{\boldsymbol{\mu}}^g(f^m)$  and can thus apply Proposition 4.18 to obtain  $(\phi, H) \in \mathcal{D}_{\boldsymbol{\mu}}^g(f)$ . The construction in the proof of that proposition yields  $\phi_t \equiv 0$  for  $1 \leq t \leq n-1$ . Therefore,  $(\phi_0, \phi_n, H) \in \mathcal{D}_{\mu_0, \mu_n}^n(f)$  and

$$\mu_0(\phi_0) + \mu_n(\phi_n) = \boldsymbol{\mu}(\phi) \leq \liminf_{m \rightarrow \infty} \boldsymbol{\mu}(\boldsymbol{\phi}^m) = \liminf_{m \rightarrow \infty} \mu_0(\phi^m) + \mu_n(\psi^m).$$

□

*Proof of Theorem 9.6.* On the strength of Proposition 9.7, the proof is analogous to the one of Theorem 5.2. □

### 9.3 Monotone Transport

The analogue of our result on left-monotone transports is somewhat degenerate: with unconstrained intermediate marginals, the corresponding coupling is the identical transport in the first  $n-1$  steps and the (one-step) left-monotone transport in the last step. The full result runs as follows.

**Theorem 9.8.** *Let  $P \in \mathcal{M}^n(\mu_0, \mu_n)$ . The following are equivalent:*

- (i)  *$P$  is a simultaneous optimizer for  $\mathbf{S}_{\mu_0, \mu_n}^n(f(X_0, X_t))$  for all smooth second-order Spence–Mirrlees functions  $f$  and  $1 \leq t \leq n$ .*

(ii)  $P$  is concentrated on a left-monotone set  $\Gamma \subset \mathbb{R}^{n+1}$  such that

$$\Gamma^{n-1} = \{(x, \dots, x) : x \in \Gamma^0\}.$$

(iii) For  $0 \leq t \leq n-1$ , we have  $P \circ (X_t)^{-1} = \mu_0$  and  $P \circ (X_t, X_n)^{-1}$  is the (one-step) left-monotone transport in  $\mathcal{M}(\mu_0, \mu_n)$ .

There exists a unique  $P \in \mathcal{M}^n(\mu_0, \mu_n)$  satisfying (i)–(iii).

*Proof.* A transport  $P$  as in (iii) exists and is unique, because the identical transport between equal marginals and the left-monotone transport in  $\mathcal{M}(\mu_0, \mu_n)$  exist and are unique; cf. Proposition 6.3. The equivalence of (ii) and (iii) follows from the same proposition and the fact that the only martingale transport from  $\mu_0$  to  $\mu_0$  is the identity.

Let  $P \in \mathcal{M}^n(\mu_0, \mu_n)$  satisfy (i). In particular,  $P$  is then an optimizer for  $\mathbf{S}_{\mu_0, \mu_n}^n(f(X_0, X_n))$ , which by Proposition 6.3 implies that  $P_{0n} = P \circ (X_0, X_n)^{-1}$  is the (one-step) left-monotone transport in  $\mathcal{M}(\mu_0, \mu_n)$ . For  $t = 1, \dots, n-1$ ,  $P$  is an optimizer for  $\mathbf{S}_{\mu_0, \mu_n}^n(-\mathbf{1}_{\{X_0 \leq a\}}|X_t - b|)$ , for all  $a, b \in \mathbb{R}$ . This implies that  $P_{0t}$  transports  $\mu_0|_{(-\infty, a]}$  to the minimal element of  $\{\theta : \mu_0|_{(-\infty, a]} \leq_c \theta \leq_{pc} \mu_n\}$  in the sense of the convex order, which is  $\theta = \mu_0|_{(-\infty, a]}$ . Therefore,  $P_{0t}$  must be the identical transport for  $t = 1, \dots, n-1$  and all but the last marginal are equal to  $\mu_0$ .

Conversely, let  $P \in \mathcal{M}^n(\mu_0, \mu_n)$  have the properties from (iii). Then,  $P$  is optimal for  $\mathbf{S}_{\mu_0, \mu_n}^n(-\mathbf{1}_{\{X_0 \leq a\}}(X_t - b)^+)$  for all  $1 \leq t \leq n$  and this can be extended to the optimality (i) for smooth second-order Spence–Mirrlees functions as in the proof of Theorem 7.12.  $\square$

## References

- [1] M. Beiglböck, A. M. G. Cox, and M. Huesmann. The geometry of multi-marginal Skorokhod embedding. *In preparation*.
- [2] M. Beiglböck, A. M. G. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *To appear in Invent. Math.*, 2014.
- [3] M. Beiglböck, A. M. G. Cox, M. Huesmann, N. Perkowski, and D. J. Prömel. Pathwise super-replication via Vovk’s outer measure. *Preprint arXiv:1504.03644v1*, 2015.
- [4] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer. Optimal and better transport plans. *J. Funct. Anal.*, 256(6):1907–1927, 2009.
- [5] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: a mass transport approach. *Finance Stoch.*, 17(3):477–501, 2013.



- [6] M. Beiglböck, P. Henry-Labordère, and N. Touzi. Monotone martingale transport plans and Skorohod embedding. *To appear in Stochastic Process. Appl.*, 2015.
- [7] M. Beiglböck, M. Huesmann, and F. Stebegg. Root to Kellerer. In *Séminaire de probabilités XLVIII*, volume 2168 of *Lecture Notes in Math.*, pages 1–12. Springer, Berlin, 2016.
- [8] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 2016.
- [9] M. Beiglböck and N. Juillet. Shadow couplings. *Preprint arXiv:1609.03340v1*, 2016.
- [10] M. Beiglböck, M. Nutz, and N. Touzi. Complete duality for martingale optimal transport on the line. *To appear in Ann. Probab.*, 2015.
- [11] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control. The Discrete-Time Case*. Academic Press, New York, 1978.
- [12] S. Biagini, B. Bouchard, C. Kardaras, and M. Nutz. Robust fundamental theorem for continuous processes. *To appear in Math. Finance*, 2014.
- [13] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.*, 25(2):823–859, 2015.
- [14] A. M. G. Cox and J. Obłój. Robust pricing and hedging of double no-touch options. *Finance Stoch.*, 15(3):573–605, 2011.
- [15] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields*, 160(1–2):391–427, 2014.
- [16] Y. Dolinsky and H. M. Soner. Martingale optimal transport in the Skorokhod space. *Stochastic Process. Appl.*, 125(10):3893–3931, 2015.
- [17] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, 24(1):312–336, 2014.
- [18] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport in general dimensions. *Preprint arXiv:1508.01806v1*, 2015.
- [19] G. Guo, X. Tan, and N. Touzi. Optimal Skorokhod embedding under finitely many marginal constraints. *SIAM J. Control Optim.*, 54(4):2174–2201, 2016.
- [20] P. Henry-Labordère, J. Obłój, P. Spoida, and N. Touzi. Maximum maximum of martingales given marginals. *Ann. Appl. Probab.*, 26(1):1–44, 2016.
- [21] P. Henry-Labordère, X. Tan, and N. Touzi. An explicit version of the one-dimensional Brenier’s theorem with full marginals constraint. *To appear in Stochastic Process. Appl.*, 2014.

- [22] P. Henry-Labordère and N. Touzi. An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.*, 20(3):635–668, 2016.
- [23] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and associated martingales, with explicit constructions*, volume 3 of *Bocconi & Springer Series*. Springer, Milan, 2011.
- [24] D. Hobson. Robust hedging of the lookback option. *Finance Stoch.*, 2(4):329–347, 1998.
- [25] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, Berlin, 2011.
- [26] D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance Stoch.*, 19(1):189–214, 2015.
- [27] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math. Finance*, 22(1):31–56, 2012.
- [28] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, Berlin, 2nd edition, 2003.
- [29] N. Juillet. Martingales associated to Peacocks using the curtain coupling. *Preprint*.
- [30] N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1823–1843, 2016.
- [31] S. Källblad, X. Tan, and N. Touzi. Optimal Skorokhod embedding given full marginals and Azéma–Yor peacocks. *To appear in Ann. Appl. Probab.*, 2015.
- [32] H. G. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67(4):399–432, 1984.
- [33] H. De March and N. Touzi. Irreducible convex paving for decomposition of multi-dimensional martingale transport plans. *Preprint arXiv:1702.08298v1*, 2017.
- [34] A. Neufeld and M. Nutz. Superreplication under volatility uncertainty for measurable claims. *Electron. J. Probab.*, 18(48):1–14, 2013.
- [35] M. Nutz and F. Stebegg. Canonical supermartingale couplings. *Preprint arXiv:1609.02867v1*, 2016.
- [36] J. Oblój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390, 2004.
- [37] H. M. Soner, N. Touzi, and J. Zhang. Dual formulation of second order target problems. *Ann. Appl. Probab.*, 23(1):308–347, 2013.

- [38] F. Stebegg. Model-independent pricing of Asian options via optimal martingale transport. *Preprint arXiv:1412.1429v1*, 2014.
- [39] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
- [40] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *Ann. Probab.*, 41(5):3201–3240, 2013.
- [41] N. Touzi. Martingale inequalities, optimal martingale transport, and robust superhedging. In *Congrès SMAI 2013*, volume 45 of *ESAIM Proc. Surveys*, pages 32–47. EDP Sci., Les Ulis, 2014.